359/7.9.70

Psychological Bulletin

LYLE H. LANIER, Editor, UNIVERSITY OF ILLINOIS
MARGARET K. HARLOW, Managing Editor

CONSULTING EDITORS

STEUART H. BRITT
Needham, Louis and Brorby, Inc., Chicago
DORWIN CARTWRIGHT
University of Michigan
FRANK A. GELDARD
University of Virginia
JAMES J. GIBSON
Cornell University
DAVID A. GRANT
University of Wisconsin
WILLIAM T. HERON
University of Minnesota
ERNEST R. HILGARD
Stanford University
WILLIAM A. HUNT
Northwestern University
JEAN WALKER MACFARLANE
University of California
DONALD G. MARQUIS
University of Michigan
JOHN T. METCALF
University of Vermont
JAMES G. MILLER
University of Chicago
NEAL E. MILLER
Yale University
HELEN PEAK
University of Michigan
ROBERT R. SEARS
Harvard University
ROBERT L. THORNDIKE
Teachers College, Columbia University

VOLUME 48, 1951

PUBLISHED BI-MONTHLY BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1515 MASSACHUSETTS AVE., N.W.,
WASHINGTON 5, D. C.

Bureau Ednl. Psy. Research
TRAINING COLLEGE

Bureau Edni. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated 7.9.70
Acc. No. J — 359

CONTENTS OF VOLUME 48

ORIGINAL CONTRIBUTIONS, GENERAL REVIEWS, DISCUSSIONS

The investigation of concept formation: W. Edgar Vinacke.... 1
Methodology in progaganda research: T. R. Vallance......... 32
A critical review of the validity and rationale of the forced-choice technique: Robert M. W. Travers............ 62
"Taste blindness" to phenylthiocarbamide and related compounds: William C. Boyd........................ 71
Johnson's review of "Work and effort": M. E. Bitterman...... 75
Ryan's "Work and effort": A reply: H. M. Johnson........... 77
A note on Louttit's review of "Individual behavior": Robert F. Creegan........................ 79
Letter to the Editor on Peters' reply to Lewis and Burke: C. J. Burke.................................... 81
A critical review of latent learning and related experiments: Donald Thistlethwaite.......................... 97
Experimental evidence for the motivating function of reward: John P. Seward.............................. 130
Comments on Razran's discussion of stimulus generalization: G. Robert Grice.............................. 150
Reply to Grice's "Comments on Razran's discussion of stimulus generalization": Gregory Razran.................. 153
A statistical consideration in psychological research: Bryan Wilkinson.............................. 156
Salivary pH as a psychophysiological variable: J. C. Gilchrist and Ernest Furchtgott......................... 193
Research with the Wechsler-Bellevue Test: 1945–1950: Albert I. Rabin and Wilson H. Guertin.................... 211
A psychophysical method employing a modification of the Müller-Urban Weights: Stephen Harrison and Margaret J. Harrison.............................. 249
Some comments on the controlling of differences in responses on the Rorschach test: Robert M. Eichler............. 257
The sensory capacities of infrahuman mammals: vision, audition, gustation: Philip Ash.......................... 289
Experimental naïveté and experiential naïveté: Richard Christie.................................... 327
The use of extended control-group designs in human relations studies: Ralph R. Canter, Jr.................... 340
The comparison of series of proportions: The problem and a new solution: R. C. Davis.......................... 348
The Freudian theory of daydreams: An operational analysis: William Seeman.......................... 369
The history of the leaderless discussion group technique: H. L. Ansbacher.......................... 383

Computation of the level of significance in the F-test: C. J. Burke........ 392
On the use of latin squares in psychology: Quinn McNemar.... 398
The genetics of schizophrenia: Reply to Pastore: Lewis A. Hurst........ 402
Correlation versus curve fitting in research on accident proneness: Reply to Maritz: Milton L. Blum and Alexander Mintz........ 413
Concerning taste blindness to PTC: Jozef Cohen and Donald P. Ogdon........ 419
Reply to Travers' "Critical review of the validity and rationale of the forced-choice technique": Donald E. Baier......... 421
Note on Travers' critical review of the forced-choice technique: Marion W. Richardson........ 435
Emotions conceptualized as intervening variables—with suggestions toward a theory of frustration: Judson S. Brown and I. E. Farber........ 465
A chi-square test for "proneness" in accident data: C. J. Burke... 496
Three rational methods for the reduction of skewness: Raymond H. Burros........ 505
Studies in social psychology in World War II: Daniel Katz.... 512

BOOK REVIEWS

Chapanis, Garner, and Morgan's Applied experimental psychology: Paul M. Fitts........ 83
Cavan, Burgess, Havighurst and Goldhamer's Personal adjustment in old age: Raymond G. Kuhlen........ 85
Burrow's The neurosis of man: Rex M. Collier........ 87
Fleming's Adolescence: Richard T. Sollenberger........ 88
Johns' Psychology in everyday living: Harry Ruja........ 89
Standen's Science is a scared cow: Karl F. Muenzinger........ 90
Guilford's Fundamental statistics in psychology and education: Wayne H. Holtzman........ 93
Stone's Annual review of psychology: Howard H. Kendler........ 159
Sargent's Social psychology: an integrative interpretation: M. Brewster Smith........ 161
Gurvitch's Sociometry in France and the United States. A Symposium: Launor F. Carter........ 163
Roback's Personality in theory and practice: Percy Black........ 164
May's The meaning of anxiety: Robert W. White........ 166
Harsh and Schrickel's Personality development and assessment: Julian B. Rotter........ 168
Newcomb's Social psychology: Glen Heathers........ 169
Allport's The individual and his religion: A. A. Roback........ 173
Munn's Handbook of psychological research on the rat: C. T. Morgan... 175
Katz's Gestalt psychology: its nature and significance: Edwin G. Boring. 177
Prothro and Teska's Psychology—a biosocial study of behavior: C. N. Rexroad........ 180
Klineberg's Tensions affecting international understanding: Dorwin Cartwright........ 181

Zahl's Blindness: Modern approaches to the unseen environment and Chevigny and Braverman's The adjustment of the blind: C. M. Louttit 182
Greenblatt, Arnot, and Solomon's Studies in lobotomy: Carney Landis . . 183
Mettler's Selective partial ablation of the frontal cortex: Frampton B. Price 185
Boring's A history of experimental psychology: C. H. Graham 260
Engle's Psychology: Its principles and applications: Frederick A. Courts 263
Guthrie and Powers' Educational psychology: Frank K. Shuttleworth 264
Anastasi and Foley's Differential psychology: Dale B. Harris 265
Hoppock's Group guidance: Principles, techniques, and evaluation: Wilbur S. Gregory 267
Plant's The envelope: Richard A. Littman 268
Edwards' Experimental design in psychological research: Irvin L. Child 269
Cochran and Cox's Experimental designs: David A. Grant 272
Harris' Some relations between vision and audition: Frank A. Geldard . . 273
Pastore's The nature-nurture controversy: John E. Anderson 273
Hurlock's Child development: S. O. Roberts 275
von Bonin's Essay on the cerebral cortex: Henry W. Nissen 276
Parten's Surveys, polls and samples: Practical procedures: Eli S. Marks 277
Freeman's Theory and practice of psychological testing: Frederick B. Davis 279
Hepner's Psychology applied to life and work: Irving R. Weschler . . . 280
Fromm-Reichmann's Principles of intensive psychotherapy: O. H. Mowrer 281
Lefever, Turrell, and Weitzel's Principles and techniques of guidance: Joseph V. Hanna 283
Kent's Mental tests in clinics for children: Helmer R. Myklebust 284
Hahn and MacLean's General clinical counseling: Ralph F. Berdie 285
Landis and Bolles' Textbook of abnormal psychology: Robert P. Fischer 286
Watkins' Hypnotherapy of war neuroses: Ralph D. Norman 287
Gibson's The perception of the visual world: Edwin G. Boring 360
Harris *et al.* Recent advances in diagnostic psychological testing: William A. Hunt 363
Wortis' Soviet psychiatry: Ivan D. London 364
von Frisch's Bees: Their vision, chemical senses, and language: F. A. Beach 365
Brown and Gilhousen's College psychology: R. C. Davis 366
Wiener's The human use of human beings: B. F. Skinner 367
Mowrer's Learning theory and personality dynamics: E. R. Hilgard and E. L. Walker 438
Völgyesi's Hypnosetherapie und psychosomatische Probleme: Hans H. Strupp 442
Rosenblueth's The transmission of nerve impulses at neuro-effector junctions and peripheral synapses: Lorrin A. Riggs 444
Gouldner's Studies in leadership; leadership and democratic action: Launor F. Carter 445

Cantril's Public opinion 1935–1946: Ross Stagner 447
Britt's Selected readings in social psychology: Kimball Young 448
Homan's The human group: M. Brewster Smith 449
Kubie's Practical and theoretical aspects of psychoanalysis: Jessie L. Miller and J. G. Miller 451
Fisher's The meaning and practice of psychotherapy: O. H. Mowrer 453
Deming's Some theory of sampling: Frederick Mosteller 453
Fryer and Henry's Handbook of applied psychology: R. L. Thorndike 455
Scheinfeld's The new you and heredity: L. M. Terman 456
Gulliksen's Theory of mental tests: Quinn McNemar 458
Stouffer *et al.* The American soldier: Adjustment during army life: Daniel Katz 512
Stouffer *et al.* The American soldier: Combat and its aftermath: Daniel Katz 512
Hovland *et al.* Experiments on mass communication: Daniel Katz 512
Stouffer *et al.* Measurement and prediction: Daniel Katz 512
Lewin's Field theory in social science: Selected theoretical papers: M. Brewster Smith 520
Festinger, Schacter, and Back's Social pressures in informal groups: Ralph R. Canter, Jr. 522
Deutsch and Collins' Interracial housing: A psychological evaluation of a social experiment: Albert Pepitone 523
Rashevsky's Mathematical biology of social behavior: C. L. Burke 525
Miller and Form's Industrial sociology: J. K. Hemphill 527
Toman's Einführung in die moderne psychologie: Karl F. Muenzinger .. 528
Porteus' The Porteus Maze Test and intelligence: D. A. Worcester ... 529
Keys *et al.* The biology of human starvation: L. P. Herrington 530
Lucas and Britt's Advertising psychology and research: Lester Guest. 532
Bloom and Broder's Problem-solving processes of college students: W. Edgar Vinacke 533
Hartley, Birch, and Hartley's Outside readings in psychology: C. F. Scofield 534
Piaget's The psychology of intelligence: W. H. Holtzman 535
Dorcus and Jones' Handbook of employee selection: Carol S. Bellows and R. M. Bellows 537
Brayfield's Readings in modern methods of counseling: W. M. Gilbert ... 538
Abt and Bellak's Projective psychology: Irving Lazar 539
Maslow and Mittelmann's Principles of abnormal psychology: P. H. Mussen 541
Horney's Neuroses and human growth: Albert Ellis 542
Redl and Wattenberg's Mental hygiene in teaching: Fred McKinney 544
Wepman's Recovery from aphasia: L. I. O'Kelly 544

MISCELLANEOUS

Books and Materials Received: 94, 190, 461, 547
Index of Subjects: 550
Index of Authors: 552

Vol. 48, No. 1 January, 1951

Psychological Bulletin

THE INVESTIGATION OF CONCEPT FORMATION[1]

W. EDGAR VINACKE
University of Hawaii

Despite the fact that concept formation has been the subject of numerous investigations, it is still poorly understood. The reasons for this lie, apparently, in two directions. First, the evolution of psychology has not gone far enough to free the treatment of concept formation from its past associations with epistemology and formal logic. Thus, terms like "abstraction" and "generalization" are still utilized—and still influence the nature of experiments—without sufficient analysis of the behavioral and genetic processes involved. Second, the data utilized in discussions of the subject are much too narrow, since they are usually drawn from limited experimental situations, and usually emphasize simple, readily quantifiable overt responses. The methods of investigation have usually been restricted in scope, and have lent themselves better to showing the results of concept formation than to revealing its nature.

It is the purpose of this article to review the research pertaining to concept formation towards the end of suggesting a broader basis for understanding it than has ordinarily been evident in experimental studies. The data available can be organized with reference to five major points. The first of these concerns the characteristics of a concept. Here one may inquire what kinds of facts must be taken into account in order to define a concept adequately. Much more than the verbal report of the subject is clearly involved, and a number of research studies have revealed internal and organizational, as well as behavioral, aspects of concepts.

The second point concerns the process (or processes) whereby an individual forms concepts. In this connection, two principal types of theories may be identified, one seeming to emphasize abstraction, the other generalization. It will be evident that neither theory is really complete, and it is likely that the two could meaningfully be combined.

---

[1] The writer is much indebted to Dr. Leslie J. Briggs for a critical and helpful reading of the manuscript.

Bureau Ednl. Psy. Research
DAVID HA L T AINING COLLEGE
Dated ....
Acc. No

It will further be shown that three different (but no doubt related) problems have been imperfectly recognized in the study of concept formation. An attempt will be made to formulate them and to relate the data to them.

The third point concerns the methodology of experiments on concept formation. As noted above, the methods employed have been surprisingly few in number; indeed, one may almost speak of established conventions in this respect. When the methodology is reviewed, it becomes evident that only two methods have commonly been used with children and three with adults.

The last two points to be taken up include a review of the experimental results with children and with adults. Research with children has mainly been directed towards a discovery of the characteristics of their concepts together with the conditions related to the development thereof, such as chronological and mental age, socio-economic status, etc. Research with adults has for the most part consisted of efforts to determine the effects of various conditions of presenting stimuli and of systematizing the behavior which is manifested in so-called concept-forming situations.

This article, therefore, attempts to organize the available research on concept formation in terms of these five major areas. In so doing it is to be hoped that a start can be made towards clarifying the problems involved in this important, but as yet inadequately investigated, area of psychological interest.

WHAT IS A CONCEPT?

None of the current definitions of a concept or of concept formation is entirely satisfactory (cf., as examples, English [9], and Smoke [58]). In view of the considerations to be raised later in the present article, it would be better not to cite any of the present definitions. Instead, we may summarize the general points which they cover. The essential features appear to be these:

1. Concepts are not direct sensory data but something resulting from the elaboration, combination, etc., thereof. This condition is easy to demonstrate experimentally by establishing a situation in which the subject cannot respond in a specified manner without data derived from previous sensory processes.
2. A corollary to the first condition, therefore, is that concepts depend upon the previous experience of the organism.
3. Concepts are responses which tie together, or link, or combine discrete sensory experiences. This condition may be demonstrated by showing that an individual responds to different stimuli in the same way.
4. It may be inferred that such ties or links are symbolic in nature; that is, the same response stands for a variety of data. In the human organism this

response is usually a word, and the word ties together different experiences with the same object, experiences with different objects somehow related to each other, the emotional responses aroused in these experiences, etc.

5. On the side of the internal processes of the organism, concepts represent selective factors. An external stimulus arouses a symbolic response, on the one hand, or a symbolic response guides perceptual response, whichever comes first. For example, a creature running along a road may evoke the response "dog"; or a boy scout may go out on a search for "birds." In experiments on concept formation, both kinds of selective mechanism are clearly revealed. Thus, an experimenter may present a whole series of objects (or a group of objects simultaneously) and ask the subject to classify them; or he may instruct the subject to find all the round objects.

In addition to these five general characteristics of concepts, various studies have suggested important aspects of how a particular concept may be organized in the mental structure of the individual.

Hayakawa (**23**) has pointed out that concepts have both *extensional* (or denotative) and *intensional* (or connotative) meanings. The extensional meaning is "something that *cannot be expressed in words* because it is that which words stand for" (**23**, p. 61). The intensional meaning is "that which is *suggested* (connotated) inside one's head" (**23**, p. 62). Whenever words are used to symbolize concepts, therefore, both extensional and intensional data may be involved.

The experiments of Reed (**52**, **53**, **54**) suggest that another way in which concepts are organized is in terms of their *consistency* or "correctness." His subjects were required to learn, by the anticipation method, the names of cards. Each card contained a series of four words, one of which fell into a category symbolized by a nonsense name. Under these conditions, Reed discovered that two kinds of concept might be formed. A *consistent* ("correct") concept was one achieved by classifying the words into groups and then attaching to the group a nonsense syllable which worked (i.e., made it possible to identify the card). On the other hand, an *inconsistent* ("incorrect") concept was one achieved by attaching a nonsense syllable to some group of unrelated words, such as the first word in each group. Both achievements must be regarded as concept formation regardless of how "objectively correct" the concept may be judged.

Although this distinction is based upon a fairly narrow experimental situation, it warrants further development. Whatever the relationship between inconsistent and intensional aspects of concepts, the two characteristics are not the same. Thus, Reed's experiments show that an inconsistent concept may nevertheless have an extensional meaning (the subject can use it to name a card correctly). Furthermore, it is quite likely that in everyday life concepts are never completely consistent but have their inconsistent components dependent upon accidental groupings of objects, inaccurate perceptions, etc.

Another way in which concepts are structured has been outlined by Welch (**68**, **70**, **71**). He has shown that they have both *horizontal* and *vertical* organization and that, in consequence, the same object (or relation) has different points of reference depending upon the other objects with which it is compared. Horizontally, objects may be classified into different categories, all of them equally inclusive. At the same time, objects may be classified vertically, into groups of varying complexity, or into more and more inclusive categories. In a given situation, the horizontal classifications are less fixed and more dynamic than the vertical. Nevertheless, even the vertical levels may be defined differently in different situations.

These important and neglected features of concepts may be clarified by means of an illustration. Figure 1 shows a collection of stimuli.

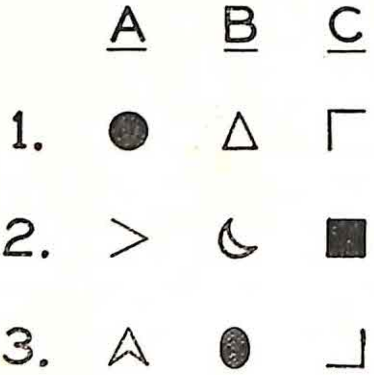

FIG. 1. A COLLECTION OF STIMULI WHICH MIGHT BE USED IN AN EXPERIMENT ON CONCEPT FORMATION.[2]

If someone were asked to select those that belong together, he would have several alternatives. Let us say that he is asked to point to the curved figures and responds by touching A-1, B-2, and B-3. This response corresponds to the *extensional* aspect of the concept. Suppose that he comments, "I dislike round objects because they're so indefinite in shape." This behavior corresponds to the *intensional* aspect of concepts.

A concept of curvedness which resulted in the selection of A-1, B-2 and B-3 would be *consistent*, whereas one which resulted in the selection of A-1, A-2, and A-3 would be *inconsistent*.

---

[2] This figure was very kindly drawn for the writer by Professor L. Scott Daniel.

In Figure 1, *horizontal* concepts would be those which resulted in the selection of A-1, B-2, and B-3, of A-2, C-1, and C-3, and of A-3, B-1, and C-2 as three classes of objects; another horizontal arrangement would be the choice of A-1, B-3, and C-2, of A-3, B-1, and B-2, and of C-1, A-2, and C-3.

A *vertical* arrangement might be as follows:

First Hierarchy —Each figure as distinct from each other figure (i.e., A-1 is a filled circle, B-1 is an empty triangle, etc.).
Second Hierarchy[3]—Group X. A-1, B-2, B-3.
Group Y. A-3, B-1, C-2.
Group Z. A-2, C-1, C-3.
Third Hierarchy[3] —Group O. A-1, A-3, B-1, B-2, B-3, C-2.
Group OO. A-2, C-1, C-3.
Fourth Hierarchy —All of the figures taken as a complex group (i.e., they are all plane figures).

Finally, note that any single figure may be related to any of the others in many different ways depending upon the concept involved. Horizontally, A-1 may be "curved," or "solid," etc. Vertically, in the second hierarchy, A-1 may be regarded as a "curved" figure, in the third hierarchy, as a "completed" figure, etc. Also consider A-3. In the third hierarchy it may be regarded as a "completed" figure from one point of view but as an "incompleted" from another.

Putting together all of these considerations, then, it appears that concepts should be placed in a broader, more dynamic context, than has been the case up to now. They must be regarded as selective mechanisms in the mental organization of the individual, tying together sensory impressions, thus aiding in the identification and classification of objects. But concepts involve more than the integration of sense impressions, against the background of which recognition occurs, for they are linked with symbolic responses which may be activated without the physical presence of external objects. That is, concepts can be given names—can be detached from specific instances, by means of a word—and used to manipulate experience over and beyond the more simple recognition function. The symbolic response, however, stands for whatever it has been linked with in the previous experience of the organism and depends upon how that past experience is organized. Thus, a symbolic response does not have a fixed, permanent meaning but represents a momentary focussing of experience upon a particular stimulus situation, combining intensional as well as extensional data, inconsistent as well as consistent relationships, and placing the stimulus in a context of greater or lesser inclusiveness.

---

[3] Note that various horizontal classifications are possible at each hierarchical level.

WHAT IS CONCEPT FORMATION?

The next question which arises has to do with the process of forming concepts. It is usual to distinguish, here, the activities of abstracting and generalizing. The former signifies the linking of one sensory experience to another so that some details are left out and others become dominant; in this sense, the concept is a response which stands for these dominant details. The latter term signifies that the dominant detail (or group of details) resulting from abstraction is used as a basis for responding similarly to the separate objects linked by abstraction and for responding to other objects similarly linked.

The two most clearly stated theories of concept formation depend upon the distinction between abstraction and generalization.

1. *Composite photograph theory.* The primary emphasis in one theory is on abstraction. As Woodworth states it, "the features common to a class of objects summate their impressions on the observer, who thus gradually acquires a picture in which the common features stand out strongly while the variable characteristics are washed out" (76, p. 801).

This view, in some form, appears to have predominated among psychologists who have written on the subject. It has been explicitly stated by Moore (43), Fisher (12), Hull (37), Kuo (38), and Stevanović (61) and is at least implicit in the interpretations of Thorndike (62) and others. The "spectator" behavior described by Heidbreder (26) belongs in the same category.

Smoke (58, 59) has criticized this theory on the grounds that the term "common features (or elements)" is misleading. He suggested that it is not specific components common to a series of stimulus patterns which summate in the concept, but relationships between them. However, he appears to be in substantial agreement with the theory.

2. *Active search theory.* By contrast, another theory emphasizes generalization rather than abstraction. Here, as Woodworth states it, "The concept is supposed to originate as a hypothesis, which *O* proceeds to test by trying it on fresh specimens of the class" (76, p. 801). This second view, therefore, assigns to the individual a more active, deliberate role than the first theory. It seeks to make of concept formation a more orderly, controlled activity than does the other view.

This theory has seldom if ever been advanced as the sole explanation of concept formation, partly, perhaps, because of the strength of the earlier theory, which has such apparently great logical consistency and sanction. Perhaps of more significance than the condition of beginning with a hypothesis is the factor of active search. Many experiments have shown that, at least in the adult, concept formation may be the result of controlled and effortful search. Thus, English (10) found that some of his subjects, in the process of abstraction began with a deliberate intention to analyze. Heidbreder (26) observed in "participant behavior" that the subject was responding with conscious and definite effort. Smoke (58) stresses the active processes of concept formation, and Reed's experiments (52, 53, 54) similarly show the influence of set and effort.

It is better to draw these two theories together rather than to make a choice between them, since experiments show that abstraction and generalization are often difficult to distinguish in the actual behavior of the organism. Under some conditions the individual may, in fact, be essentially a passive recipient of sensory impressions which gradually summate into the concept. Under other conditions it may be that an individual proceeds by establishing an hypothesis and then deliberately checking it against instances. More than likely, they are mutually complementary approaches to a situation and occur as interrelated processes. This conclusion is especially warranted with respect to concept formation in adults. It cannot be assumed that the adult subject is really acquiring a "new" concept; instead, he is learning to apply concepts (or aspects of concepts) already in his possession in some unusual or systematic fashion; i.e., he is solving a problem rather than learning a concept, in contrast to a small child who is more likely to be learning something new. In this sense, therefore, both abstraction and generalization, for the adult, are far more conscious, controlled, motivated activities than they seem to be in the child. Hence, for the adult, the distinction between abstraction and generalization breaks down. One cannot say that, for the adult, the sequence is perception-abstraction-generalization except from the logical standpoint. In a child, on the other hand, confronted with an entirely new realm of experience, such a sequence appears to be more likely.

Heidbreder's findings with respect to a complex concept formation situation (26) bear out the view that processes of both kinds operate at the adult level. She shows that concept formation involves, on the one hand, "analysis" and on the other hand, "synthesis." The subject both seeks and obtains information about the situation and also organizes his activities in relation to it. "Participant behavior" and "spectator behavior" do not occur as mutually exclusive methods of approach, although one or the other may at times predominate. They both contribute to the end result, the formation of a concept.

Problems in the Study of Concept Formation

These preliminary considerations lead to a recognition that several different problems are actually involved in the investigation of concept formation and that investigators have not clearly distinguished between them. These fundamental problems may now be stated. For convenience, identifying names are given to them.

Problem 1. Ability to conceptualize. This question may be stated explicitly in this way: How can one explain and describe the development in the child of the ability to form and use a concept? This problem is clearly a genetic problem and should be recognized as such. It cannot be answered, in all probability, by studying adults and then extrapolating from their behavior to early stages of child development. In any case, this problem is a general one and represents an effort to trace, with age, the unfolding and elaboration of a general function in the behavior of the individual together with the conditions which influence that development.

Problem 2. Acquisition of concepts, or repertory. This question may be formulated as follows: What concepts—or patterns of concepts—characterize various stages in the development of the child's thinking and acting? This is, so to speak, an educational problem. It may be that it is inextricably bound up with Problem 1, but this writer, at least, is not prepared at the present time to regard it as the same thing. The reason lies in the fact that the emphasis in Problem 2 is more specific than that in Problem 1; i.e., the concern in Problem 2 is with the particular concepts which the child possesses and with the way he utilizes them. Hence, a number of additional questions arise pertaining to the order in which different concepts are acquired, how complex they are, how many concepts the average child of a given age possesses, etc. Clearly, this problem, like Problem 1, must be answered by tracing the learning process in children, although related questions have meaning also in the study of adults.

Problem 3. Achieving a specific concept. This question may be stated in this manner: How does the individual go about attaining a particular concept? What behavior is manifested, and what conditions influence that behavior? Although this third question is pertinent to children, too, it has primarily been investigated at the adult level. In fact, when adult subjects are used it appears to be inevitable that emphasis is placed on this problem, because the adult has already developed an ability to conceptualize and has already acquired an enormous repertory of concepts. It is probable, therefore, that the adult does not typically acquire new concepts so much as he applies concepts which he already possesses—or learns new variations, hierarchies, etc., of those concepts.[4]

EXPERIMENTAL APPROACH: METHODOLOGY

In attempting to subject the problems of concept formation to standardized, objective conditions of observation, investigators have exercised a considerable degree of ingenuity. The stimuli which have been devised are clever, interesting, and novel; several excellent series of studies have been conducted; nevertheless, most of the experiments have conformed to a very few relatively simple patterns. In fact, only three basic methods have been used with adult subjects and two with children.[5] Some of the studies have employed two or three of these methods in such a manner that the distinctions to be drawn below are somewhat arbitrary.

---

[4] Cf. Hebb (25), who goes so far as to say that there are two kinds of learning, one characteristic of the child, the other of the adult.

[5] The emphasis here is on laboratory research. The methodology might be extended if one included relevant data from other fields such as intelligence testing or therapeutic situations.

Experiments with Children

The interview-questionnaire method. This approach has been employed mainly with reference to Problem 2 formulated above. Typically, the investigator works out a procedure for obtaining from the child his viewpoint or judgment about the nature of reality, morality, causal relations, life, death, etc. Piaget (**49, 50, 51**) is the best-known exemplar of this method (he calls it a "clinical method"). The procedure is essentially that of questioning the child in an objective manner about natural and psychological phenomena, striving to obtain from the child his own interpretations thereof. Obviously, there are many pitfalls in this approach, but Piaget has been patient and careful in his efforts to objectify his questions and his interpretations of the answers.

Attempts to verify Piaget's findings have modified the method in the direction of greater objective standardization, Thus, the questions have been administered in questionnaire form, or each child has been asked the same questions in the same way (e.g., **3, 6, 7, 46, 56, 57**). Sometimes the investigator has actually demonstrated natural phenomena, then asked the child to explain them (**7, 46**).

The performance method. The second kind of approach has been used primarily to investigate Problem 1. Here, the investigator places the child in a situation which involves the learning or use of one or more concepts and observes the child's behavior. Thus, a child is taught that if he opens a box on the front of which is a triangle he will find a piece of candy. He is then presented with two boxes, one of which is marked with a triangle, the other with a square, to see if he can discriminate between them (**15, 16, 44, 65, 66, 67, 68**).

This second method has distinct limitations from the standpoint of interpreting the child's behavior. Thus, even if a child responds correctly (that is, appears to act in terms of the "concept") it is difficult to be sure that the response is not a simple conditioned response. In the training series the triangle→candy relationship is set up; in the discrimination series there may be no reason to suppose that the response is any more complicated than that (i.e., the square might simply be ignored). Care must be exercised, therefore, not to assume that the child's behavior is more complex than it really is. This difficulty can at least partially be reduced by emphasis on adequate experimental controls (such as by varying the stimuli) and by restricting interpretation as much as possible to the actual behavior of the child. With children old enough to verbalize, introspective data might have considerable value in supplementing the behavioral data.

Experiments with Adults

The introspective method. The essential characteristic of this method is that the subject is exposed to a series of stimuli and asked to analyze and report the mental processes resulting from the stimuli. The emphasis is upon the acts or processes in the individual which are not directly accessible to others. The method at the present time is mainly of historical interest. Early investigations of concept formation made use of it (10, 12, 43). Thus, a series of drawings might be exposed tachistoscopically until the subject believes that he has discovered something important about the stimuli. The exposures are stopped while the subject reports his mental processes. Experiments in which the emphasis has been on behavioral data have also included introspective reports (e.g., 26, 52, 53, 54).

It might be observed that valuable data can be obtained through introspective studies and that this method should not be abandoned too hastily. For example, studies of the intensional and inconsistent aspects of concepts might well utilize it.

The learning method. Here, the subject is placed in a situation which requires that he learn some clue or principle by means of which he can identify the members of a class. Typically, the conditions have resembled a memorizing experiment. Thus, the subject is first taught the names (usually, nonsense syllables) of a series of stimuli, say by the anticipation method. He is then shown a second series of stimuli, each of which has something in common with the first series, to see if he can correctly name them (26, 28, 37, 52, 61).

The problem-solving method. The third approach is similar to the learning method except that the subject does not need to go through a preliminary learning period. He is usually presented with a collection of stimuli, either simultaneously or successively, and asked to determine what they have in common or to classify them in some way. Thus, the subject may be told that a series of drawings are all instances of the same concept; he is to examine them and explain or define the concept (58). Another variation is to present the subject with a collection of objects, drawings, or cards and ask him to select those that belong together (2, 19, 20, 21, 22, 34).

CONCEPT FORMATION IN CHILDREN

At the present time it would be presumptuous to suppose that we possess enough knowledge to present a complete and accurate account of concept formation in children. There is a considerable and growing body of data dealing with that and other aspects of the intellectual development of children, but so many questions regarding interrelationships and attendant conditions remain to be answered that it would be fruitless to attempt to draw all the data together. We shall, then, concentrate on the investigations which are directly pertinent to concept formation. These fall naturally into two major groups, those which purport to specify the characteristics of children's concepts and those which attempt to define the conditions related to the formation of concepts, notably age. In these studies, the three problems set forth earlier are not clearly differentiated. Thus, it is not always easy to tell whether the results are of general importance with respect to the development of the ability to use concepts ("conceptualizing") or merely with respect to how a particular concept is achieved and used by the child.

Characteristics of Children's Concepts

The work of Piaget (**49, 50, 51**) set the stage for a host of subsequent investigations. Using the method briefly described above, he obtained a wealth of provocative material relating to how the child conceives of himself and his surroundings. Perhaps the two major features of Piaget's interpretation of his results are (1) that the child's thinking is essentially different from that of the adult and (2) that the child's concepts pass through a series of rather clearly defined stages each of which is typical of a particular age-group.

According to Piaget, the child's thinking is characterized by *realism*, *animism*, and *artificialism*. By realism, Piaget means "a spontaneous and immediate tendency to confuse the sign and the thing signified, internal and external, and the psychical and physical" (**49**, p. 124). In other words, the child does not distinguish between his own experience and external reality. There are four stages in the evolution of realism, as follows (**49**, p. 126):

1. Absolute realism: "during which no attempt is made to distinguish the instruments of thought and where objects alone appear to exist."
2. Immediate realism: "during which the instruments of thought are distinguished from the things but are situated in the things."
3. Mediate realism: "during which the instruments of thought are still regarded as a kind of things and are situated both in the body and in the surrounding air."
4. Subjectivism or relativism: "during which the instruments of thought are situated within ourselves."

In general, the child is alleged to pass from a confusion of self and world to a distinction between them. For example, Piaget gives protocols relating to the notion of thought, the nature of names, and the nature of dreams. The child first confuses thought with the object thought about and only much later reaches the point of recognizing the difference between the thing, or the word, and the thought. With respect to names, the child passes from a stage of assuming that the names of things are in the things themselves, to a stage of recognizing that names are in the voice, the head, or the thought itself. Dreams are first regarded as coming from outside and remaining external, then arising in ourselves but external to us, and finally are conceived to be internal in origin and to remain internal.

Piaget means by animism in children's thinking, "the tendency to regard objects as living and endowed with will" (**49**, p. 170). He postulates four stages of development in this dimension:

1. At first, the child believes that everything active is conscious. "Thus a stone may feel nothing, but if it is moved, it will feel it" (**49**, p. 174).
2. A second stage is one in which consciousness is attributed to things that can move, such as the sun and moon, clouds, rivers, bicycles, etc.
3. In the third stage, consciousness is attributed to things that can move of their own accord. Thus, a distinction is drawn between movement exhibited by the object itself and that introduced by an external agent; the sun and wind are conscious, but not the bicycle.
4. Finally, consciousness is restricted to the animal world.

A third general characteristic is artificialism: "regarding things as the product of human creation, rather than in attributing creative activity to the things themselves" (**49**, p. 253). Thus, for example, the origin of the sun may first be explained in terms of some human agency (e.g., it was lighted with a match), and then by intermediate stages the child's thinking progresses to the point where it is recognized that human activity has nothing to do with the phenomenon.

In presenting his findings with respect to many different concepts, Piaget defines the successive stages through which that particular concept passes, and links them with age. For example, he alleges that the four stages of animism occur, on the average, at ages 4 to 6, 6 to 7, 8 to 10, and 11 or older. In the case of physical causality (**50**), he finds that some seventeen types of explanation are advanced, falling into three main stages. In the first, up to about ages 5 to 6, the child's explanations are "precausal," representing a confusion between psychological and mechanical factors. In the second, at about ages 7 to 8, children explain natural phenomena in ways which are still essentially "precausal," but the magical types tend to disappear. Finally, in a third stage, which is completed at about ages 11 to 12, more rational forms become dominant.

Thus, Piaget has found in many different areas the development in children of concepts which pass from diffuse, prelogical, subjective forms to more differentiated, logical, and objective forms. He claims that general stages are characteristic of particular ages, and that concepts of a particular kind (e.g., of "life") evolve in a definite manner.

Nagy (**45**) obtained results similar to Piaget's from a study of children's theories of death. She obtained data consisting of compositions, drawings, and discussion from 378 children. Three stages were found, as follows: (a) denial of death as a regular and final process; i.e., everything including the dead and lifeless is considered to be alive—the animism described by Piaget (ages 3 to 5); (b) death is personified, considered a person—essentially Piaget's artificialism (ages 5 to 9); and (c) death is understood as a process and as inevitable (ages 9 and older).

Some rather serious criticisms of Piaget's views have been given. Hazlitt (**24**) has suggested that his claims rest upon two unwarranted assumptions, namely: (a) that thinking can be identified with ability to express ideas verbally; and (b) that the content of thought is irrelevant—that is, if a child cannot understand or state a relation in regard to one subject matter, he cannot do so in regard to another. She herself conducted experiments designed to have the child behave in terms of a concept rather than merely to verbalize it. She concluded that children display a grasp of relations very early and that the only limitation is lack of experience. In general, an assumption that there is a qualitative difference between adult and childish thinking is not warranted. Deutsche (**7**) has done a very careful study of children's concepts of causal relations, as a result of which she criticizes Piaget's classification of answers into seventeen types. In the first place, she found that the definition of the types is so vague and indefinite that it is difficult to classify answers. In the second place, the types are of extremely unequal value, some being too inclusive, some extremely narrow, although Piaget tends to treat them as though all were characteristic of children's concepts. Finally, it is a matter of personal judgment as to what a child really means by his explanations. Many criticisms brought to bear against Piaget rest upon his methods, which are indeed difficult to accept fully; for example, one never knows how many subjects he is relying upon in the formulation of his conclusions and one never knows how many children were asked the same questions. Other criticisms depend in part upon Piaget's terminology, although he himself admits the fallibility of such words as "animism" and "realism."

A large number of studies, however, have been carried out with more standardized procedures than those of Piaget, some of which have confirmed his results, others of which have disagreed with them. Grigsby (**18**) investigated developmental trends in concepts of time, space, cause, part-whole relations, discordance, and number, using children between the ages of 2 years 8 months and 6 years 4 months. In each of these kinds of concepts she found gradations, although these were more related to mental age than to chronological age. Specifically, with respect to causal relations, she found good agreement with Piaget's seventeen types.

This is in marked contrast to the results of Deutsche (**7**), already mentioned, whose experiment is in many ways a model of research in this field. She developed two tests, both concerned with causal relations, the answers for which were written. In the first test there were eleven demonstrations of physical phenomena which the child was asked to explain. For example, the experimenter placed a jar over a lighted candle and then asked, "Why does the candle go out?" In the second test there were twelve questions, such as, "What makes the wind blow?" and "What makes the snow?" She used 732 subjects (thereby meeting at least one criticism of Piaget) distributed rather equally over the age range 8 to 16—both boys and girls. They were also distributed over Grades 3 to 8 and corresponded well with the percentage of the male population of Minneapolis falling into seven occupational groups. I.Q.'s ranged between the intervals of 60–69 and 140 and above. The enormous number of answers obtained were carefully and objectively evaluated in several ways by several judges (in one instance by thirteen judges, in another by three). The following are her major conclusions pertinent to the characteristics of children's concepts of causal relations:

> Classification of all the answers to Piaget's seventeen types of causal thinking revealed that only four types of thinking (phenomenistic, dynamic, mechanical, and logical deduction[6]) are found in large enough frequency to warrant further analysis.
>
> No evidence was found that children's reasoning develops by stages. Both quantified and qualitative analyses show a gradual progression in answers with advancing age, and all kinds of answers are found spread widely over the age range.
>
> Evidence was found of specificity in children's causal thinking as opposed to a general level of thinking. The child responds to different questions with answers differing in qualitative as well as quantified value.

However, Deutsche's study itself has certain limitations. Perhaps the most important of these is the nature of her questions. They were not as varied as those of Piaget. In fact, the absence of some types of answers in her data might be accounted for by the difference in the kinds of question which she asked. Another limitation has to do with the ages of her subjects, the youngest of whom was eight. By this age, according to Piaget, some of the precausal types of answer have largely disappeared. As a check on this point, Deutsche administered the tests to thirteen kindergarten children (five years old). Despite the fact that these data are not presented in the same way as those for the other subjects, they seem to agree fairly well with Piaget's formulation. That is, the younger subjects are much lower in total score, and the precausal types of answer appear more frequently. Nevertheless, even at the five-year level, the same general range in type of answer is evident as in the older levels.

Russell and Dennis (**56**) devised a standardized procedure for investigating animism in children's concepts. It consisted of a questionnaire relating to twenty objects, some physically present, some visible outside the room, and others presented verbally. They tested 385 children ranging in age from 3 years to 15 years 6 months. Criteria were set up whereby the answers could be classified into five categories, the four stages previously formulated by Piaget and a "No Concept Stage." This method was then used to check Piaget's contention that these categories of response are characteristic of definite ages (57). The subjects numbered 774, ranging from Grade 1 to Grade 8. It was found that 98.5 per cent of the subjects could readily be classified into one of the stages. Curves based on the data showed increasing modes with increasing mental age, from which result Russell concludes that it is probable that individuals pass sequentially through the stages. Nevertheless, as Deutsche had found, for causal concepts all the stages occurred at every age, rather than being distinct from each other as Piaget had described them.

---

[6] *Phenomenistic:* two facts given together in perception, no relation except contiguity in time and space; *dynamic:* no animism, but still sees in objects forces capable of explaining their activity and movements; *mechanical:* explanation by contact and transference of movement; *logical deduction:* explanation by the principle of sufficient reason.

Dennis (6) also applied Piaget's questions to Zuni and Navaho children. He found that animism, realism, and artificialism are also typical of their concepts. Bruce (3) similarly investigated the occurrence of animism in Negro compared to white children. She found Piaget's four stages, but she believes that the concept of "alive" is a more meaningful way to interpret children's beliefs than to refer them to "animism." Like other investigators, she found that all four stages occur at all age levels.

The experiment by Oakes (46) tends to agree with Deutsche. His method was patterned closely after Piaget, and he attempted to obtain age sampling at younger ages than did Deutsche. Thus, he had 77 kindergarten subjects, from age 4 years 10 months to 7 years 4 months, in addition to smaller numbers of children in Grades 2, 4, and 6. He used fifteen kinds of questions ("How were the hills made?") and seventeen simple experiments similar to those of Deutsche. He found, as Deutsche did, that Piaget's seventeen types of explanation are not usable; instead, he classified the answers into three major categories, physical or materialistic, non-physical or non-materialistic, and failure to explain. For both types of tests there was a marked increase in the incidence of physical explanations with age, accompanied by a corresponding decrease in the other categories. In his results, too, all types of answer were given by all the age groups, and each subject gave a wide variety of explanations. Hence, he found no evidence for Piaget's contention that definite stages characterize definite ages. "The types of answers given by these subjects were influenced by the nature of the problem . . . " (46, p. 93). The great majority of the answers were naturalistic, which contrasts with Piaget's emphasis on various kinds of magical, illogical, etc., explanations.

Finally, it should be noted that experiments of very different kinds have also cast doubt upon Piaget's view of age-stages in concept formation. For example, Welch and Long (42, 70) tested children between the ages of 42 and 83 months with tasks which required the use of concepts previously learned. Although the number of subjects who could pass the tests increased with age, there was no orderly progression in this respect.

One general conclusion to be drawn from these studies is that children's concepts seem to change with increasing age but that the change is more in the nature of a gradual progression (towards greater adequacy or informed understanding, perhaps) than of definite stages. This change does not occur at the same rate for all children. Indeed, some may never achieve the more advanced kinds of explanation at all, since even adults display concepts similar to those of children. Further, no child at a particular age gives consistent responses of one type or another but instead may give many different types, depending upon the situation.

Another conclusion to be stated tentatively is that the apparent difference between children's concepts and those of adults is less a matter of kind than of degree. Although animism, artificialism, and realism—to retain Piaget's terms—are found relatively more often in children, these characteristics are not absent by any means from the thinking of adults. The reverse is also true, since the supposedly more advanced, more logical, more correct characteristics occur in the concepts of children. As will further be shown in the following section, it is probable that Hazlitt's observation is correct: the greatest factor in the difference between concepts of children and adults is that the former have less experience and knowledge (**24**).

A further observation which follows from this has been emphasized by Wolff (**75**). It is necessary, in order to understand the concepts of children, to judge from the child's standpoint, not that of the adult. Thus, what strikes the adult as "illogical" may not be so from the child's outlook, which is merely based upon less, or different, experience.

Conditions Related to Concept Formation in Children

In this second area of research much remains to be done—for example, in terms of relating what is known about development specifically to concept formation. We shall merely indicate some of the possible conditions which have emerged in experiments dealing with concept formation as such.

Age. Welch and Long (**42, 68, 70, 71**) have shown that the conceptualizing ability of children seems to develop from simple to more complex levels. A pre-abstract period leads gradually to an ability to grasp first-hierarchy concepts, such as that "men" and "women" are all "people." This period begins at about the twenty-sixth month and is followed, by about the middle of the fourth year, by the ability to grasp second-hierarchy concepts ("potatoes" are "vegetables," "apples" are "fruit," and both "vegetables" and "fruit" are "food"). In turn, the child may develop an ability to grasp concepts of still higher hierarchies in later years. It should be noted, in relation to the preceding section, that at least after the age of eight years children can deal with concepts at any of the lower hierarchical levels (**42**).

Reichard, Schneider, and Rapaport (**55**) have used sorting tests to study the conceptual abilities of children. They suggest that there are three levels of development: (1) *a concretistic level*, where classification tends to be made on the basis of non-essential incidental features of the objects; (2) *a functional level*, where classification is on the basis of use or value, etc.; and (3) *a conceptual level*, where the child classifies more nearly on the basis of abstract properties or relations. The concretistic performance was more typical of the youngest subjects (up to 5 or 6 years), and the functional performance more of older children (up to 8 or 10 years).[7]

Still another change in conceptualizing with age is suggested by the work of Colby and Robertson (**5**), who tested three age-groups (3 to 6, 6 to 7, 7 to 9) with colored forms and retested them after one year. There was a progressive shift towards a dominance of form as the basis for classification. Welch (**66, 68**) reports a similar result, although he concludes that form tends to be dominant below the age of 3 years, with a shift to color between 3 to 6, and a return to form again after age 6.[8]

Other efforts to trace conceptualizing in relation to age may be summarized as follows:

Triangularity and other forms (**15, 16, 44**): concept can be established at 15 months to 2 years.

Roundness (**40**): concept established as early as age 3.

Magnitude (**36, 63, 65, 66, 67, 69**): simple concepts of size can be established quite early, perhaps at the fourteenth month, but "middlesizeness" comes later, probably not much earlier, on the average, than the third year.

Contradictory relations (**8**): this situation is one in which children are asked to discriminate the larger of two equal ring-segments so placed that one gives the illusion of being larger than the other; the larger is then placed so that it appears smaller. Recognition of the contradiction was found to appear between ages 3 and 4, the same time as the concepts "big" and "little."

Time (**1, 13, 14**): concepts in this area undergo long and continuous development. General divisions of time are grasped early, "morning" or "afternoon" by 4 years, what day it is by 5 years, the conventional time-system by the sixth grade in school. Still other time concepts are grasped later, with adults about equal, on the average, to ninth- or tenth-graders.

Cause-effect relations (**39**): grasped by the eighth or ninth year but probably not before.

Social concepts (**47**): there is a steady increase from grade to grade although concepts in different problem areas differ. Recognition of social problems is not great before the sixth or seventh grade.

---

[7] See below for discussion of concrete and conceptual attitudes in adults.

[8] Apparently, form also tends to be dominant over color in adults, as would be expected from the above (see **31**).

Intelligence. It is probably safe to say that psychologists have assumed that intelligence and concept formation are related, although they have not yet worked out the relationship explicitly. There are two assumptions, really: (1) that one of the variables of intelligence is the ability to form and use concepts, and (2) that part, at least, of the reason why mental age increases during the period of growth is that the ability to conceptualize increases.

These assumptions seem reasonable, yet relatively few efforts have been made to verify them. What this amounts to, basically, is comparing the performance of children of high and low intelligence, estimated as best one can in situations which involve the formation and/or use of concepts. When this is done, some unexpected results may occur which force us to raise additional questions. Thus, Deutsche (7) found that scores on her tests of causal relations had only a low relation to intelligence test scores (Kuhlmann-Anderson). The studies of Long and Welch (42), as well as others, indicate that chronological age is at least as important as mental age.

Training or experience. This condition further complicates the relationship between intelligence and concept formation. Studies such as those of Deutsche (7), Oakes (46), and Ordan (47), suggest that children differ in their concepts and use of concepts as much because of variations in experience as in variations in intelligence. Investigations of very young children, such as those of Gellermann (15, 16) and Welch and Long (73), indicate that at a very early age children are capable of learning concepts and can, in fact, be taught them. Peterson (48), in a study of generalizing ability involving the principle of the lever, found that grade in school had a closer relation to performance than either age or measured intelligence. In Deutsche's investigation (7) the correlation between school grade and scores on her concept test (combined forms) was .445 (*PE* .067), whereas the correlation with I.Q. was only .182 (*PE* .098).

Socio-economic status. Both Deutsche (7) and Ordan (47) report that socio-economic status has only a very low relation to scores on concept tests.

Vocabulary. Both Deutsche (7) and Ordan (47) report a substantial relation between this factor and concept scores. In the former study the correlation between number of words used and scores was .48 (*PE* .06); in the latter the correlations between vocabulary and various areas[9] of social concepts were of about the same magnitude.

In summary, it may be said that the genetic aspects of concept formation have been very inadequately studied up to the present time. The most thoroughly investigated aspect of the problem appears to be that which relates to our Problem 2, that is, the acquisition of concepts. Problem 1, which concerns the more general question of "conceptualizing," remains comparatively untouched. Perhaps the most interesting feature of this problem is the apparently low correlation with intelligence test scores and the relatively high correlation with training and experience, of which one important component seems to be vocabulary.

---

[9] Crime, economics, government, war-peace, etc.

EXPERIMENTS WITH ADULTS

As noted previously, research at the adult level has primarily been concerned with our Problem 3, that is to say, less with concept formation in general (conceptualizing), and more with a situation in which a particular concept or group of concepts is achieved (that is, isolated from or applied to a particular set of stimuli). Although we distinguished between learning and problem-solving situations, the latter aspect is the predominant one. Although experimenters have striven to set up conditions in which the subject is confronted with new experiences out of which new concepts are supposed to emerge, it appears to the present writer unlikely that this can be done except in the case of children. Thus, with adults it seems probable that the main emphasis should be on such problems as the behavior of the subject, what concepts he uses, and how he uses them. In the discussion to follow it will be assumed that the subject is fundamentally using his past experience rather than acquiring new experience.[10] The data obtained from experiments may conveniently be considered under two principal headings, the conditions under which the material is presented, and the behavior occurring in concept formation situations.

Conditions of Presentation

The many excellent experiments in this field have contributed much to an understanding of efficiency in situations where concepts are to be used or experience is to be generalized.

Method of presenting stimuli. One of the earliest objective experiments, the famous one of Hull (**37**), was an effort to determine the factors which promote efficiency in this kind of problem-solving situation.

The method used by Hull has served as a prototype for many subsequent experiments. He prepared twelve series of stimuli each of which contained a common element or concept. These were presented successively in an exposure apparatus. With each exposure the experimenter pronounced the nonsense syllable which served as a name for the concept. In succeeding series the subject tried to anticipate the correct name of the stimulus but was corrected by the experimenter if he failed to assign the correct name to it. The ingenious stimuli used for this procedure were Chinese characters built around radicals (the common element, or concept). In this way Hull compared, for example, two orders of presenting instances of the material each of which contains a "common element." The first order is from simple-to-complex, the second from complex-to-simple. If the time spent on each instance is the same there is no apparent difference between the two orders, "but if each individual experience in the evolution series is continued until the reaction to it is just perfected before passing to the next, there is a distinct advantage in favor of the simple-to-complex method." In general, simple instances are more efficient because it takes less time to form the concept from them (i.e., extract the common element).

---

[10] This assumption does not mean that it might not be possible to study the acquisition and conceptualizing of experience in the adult; so far, however, experiments have not convincingly done so.

Context and critical features. Hull also studied the effect of emphasizing the concept (the common element) in various ways, in contrast to presenting only the total characters in which the concept was embedded. Thus, giving the common element outright is no more efficient than extracting it from a series of concrete instances. In the former situation, however, the subject can better verbalize the concept, although this is obviously not necessarily a true indication of functional value, since the subject did not show any better performance. If the common element is emphasized *in situ* by outlining it in a different color (red), it is easier to generalize it. So far as familiarity is concerned, it was found that twice as frequent exposure to half the series of instances is less efficient than moderate exposure to all the members of the series. Finally, the most efficient method of presentation was one which combined examples of the concept ("naked" common element) with concrete examples of it embedded in various contexts.

The elaborate and lengthy series of experiments performed by Heidbreder (**28** through **35**) shows other conditions of this kind. Her general procedure is similar to that of Hull; that is, it was an adaptation of the familiar memory situation in which the subject learns the nonsense syllable names for various categories of stimuli by the anticipation method. In carrying much further the principles suggested by Hull's experiments, she has found that concepts are attained more easily when the critical features (the features which identify the concept) are readily accessible perceptually and are more "thinglike" (**31**, **32**). For example, concepts of objects were more quickly attained than concepts of forms, and attainment was more rapid, in general, when non-critical features occurred only once but critical features occurred repeatedly. Concepts are also attained more easily when the context provides strong perceptual support for the critical features. In general, the more thing-like the instances the more readily is the concept attained, and the more the context supports a particular concept where several are possible, the more rapidly will the supported concept be attained. Thus, varying the context from experiment to experiment resulted in marked changes in the order of attaining concepts. (This point is further elaborated below.) Another line of evidence comes from the nature of the task. When the subject is permitted to participate perceptually more directly in the experiment, concepts are attained more readily. For example, when subjects sorted cards instead of observing the instances in the exposure apparatus, the concepts were more rapidly achieved. The use of verbal instances of the concepts instead of drawings yielded essentially similar results except that the order of attainment seemed to depend not upon the thing-character of the instance but upon "semantic efficiency."

Positive and negative instances. Kuo (**38**) used a method similar to that of Hull and found that the more negative or misleading characters there were in the series, the less accurate the conclusion. Smoke (**58**) used geometrical designs as stimuli. Each series was constructed according to a certain principle, the discovery and definition of which was the subject's task. In comparing performance on series containing both positive and negative instances of the principle with those in which only positive instances appeared, Smoke found no difference in efficiency. Actually, the situation may be more complex, for some subjects seemed to be aided by the negative examples whereas others did not profit from them. In a later study, Smoke (**59**) confirmed this result but felt obliged to qualify the conclusion to the effect that the presence of negative instances may serve as checks in the learning process and thus are of assistance, especially where difficult concepts are involved.

Set. One of the experiments by Reed (**52**) reveals the influence of set. The task of the subject was to learn the nonsense names of forty-two cards on each of which were written four common words. One word on each card belonged in one of six categories represented by a nonsense syllable. To study set, Reed instructed one group of subjects merely to learn the name of each card; a second group were to learn the name and secure the meaning and were also given information about the nature of a concept. The latter group, characterized by a specific set, showed a greatly superior performance. Reed suggests that this economy occurs because the latter group are more likely to keep the goal (learning the concepts) in mind and participate more effectively in the problem situation by searching for possibilities of grouping, testing hypotheses, etc. Ewart and Lambert (**11**) have also reported results of this kind.

Length of series. In another experiment Reed (**53**) examined the effect of varying the amount of material. Within the scope of his experiment longer series were more efficient than short series; that is, the amount of work required, though absolutely more, was relatively less for longer series. The proportion of consistent concepts achieved also increased with length of series.

8-359

Complexity of stimuli. In another experiment by Reed (**54**) the situation was the same, but the number of words on each card was increased and confusing categories were introduced. The latter, however, only applied to a few of the total possible number of instances. As might be expected, the effort required to form concepts increased directly with increased complexity. The most striking effect, meriting further research, was that the proportion of consistent or correct concepts decreased markedly with an increase in complexity, accompanied by a corresponding increase in the proportion of inconsistent or incorrect concepts. As Reed states it, "as the complexity of the stimuli is increased there is a definite trend to shift from logical to illogical learning, or to base concepts on such factors as the primacy, and frequency, and sensory similarity of contiguous stimuli."

Mutual interference. Gengerelli (**17**) studied another interesting phenomenon. He had one group evolve concepts from meaningless drawings which were named by a letter, then presented them with the same material in which the concepts were named by a different letter. A control group carried out the same procedure, except that the drawings were changed somewhat in the second stage. The experimental group showed more interference than the control group, thus indicating the difficulty of shifting the meaning of a concept from one situation to another.

From this survey of experiments it may be said that if concept formation is defined as the process of discovering an identifying detail, relation, or principle by means of which a series or collection of instances can be classified,[11] as it may be perfectly proper to do in keeping with our Problem 3, then we know to a considerable extent what will happen under various conditions of presenting the material. The major findings are as follows:

1. The order of the instances does not particularly matter if an equal amount of time is allowed to each.
2. Abstract instances alternated with concrete examples is the most efficient method.
3. Negative instances are not necessary but may be included to serve as an occasional check on accuracy.
4. The subject who is well instructed in his task will do better than the subject who is simply a passive recipient of the material.
5. Consistency is promoted if the series is long rather than short, and simple rather than complex.

[11] Dr. Leslie Briggs has, in fact, proposed to the writer this definition for concept formation: "The process through which the person manipulates and classifies the essential common features embedded in a number of complex stimulus situations in order to arrive at a symbol which represents the common elements in the total range of stimuli." This definition might well be adopted for our Problem 3.

6. A given context is likely to favor one concept over another, and the readiness of attainment will depend upon how accessible the critical features are to perception.

It is evident that much work remains to be done to ascertain how general these conclusions are. At present they must be regarded as tentative statements applicable to the kind of experimental situation upon which they are based.

Behavior in Concept Formation

Experimental studies have revealed a number of important phenomena which occur in the formation and use of concepts. To some extent, accordingly, it is possible to describe what happens when the individual is placed in a situation corresponding to the experimental conditions.

Gradual attainment. Several of the experiments which attempted to break down the successive phases of discovering the concept, mainly through the introspective method (**4**, **10**, **26**, **37**, **53**, **61**), have revealed that concepts are typically evolved gradually by the subject. In general, the subject is exposed to the material and little by little begins to interpret it—perhaps by perceiving in it familiar objects or associating aspects of it with his past experience; perhaps by deliberately attempting to analyze it, comparing one instance in the experimental situation with another, etc. In time, some features begin to stand out, others to "disintegrate" (**61**), until the subject is able to formulate a principle which will work. He may, of course, fail to develop an adequate concept.

Attainment without awareness. Hull (**37**) and Smoke (**53**) both found that an individual may be able to act in a manner showing that he can act in terms of the concept without knowing what it is. That is, he can successfully identify specimens of the class without being able to verbalize the concept.

Attainment of consistent and inconsistent concepts. As previously noted, Reed (**53**) found that subjects may develop inconsistent as well as consistent concepts. In general, the development of a consistent or inconsistent concept results not from different factors but from different manifestations of the same factors. These factors are set, identical perceptual elements in the concept-symbol and the instances, frequency of occurrence of key-items belonging in a given class, contiguity between an instance and the class-symbol (e.g., position factors), and generalization. Consistent concepts are favored by a specific set directed towards group classification, whereas inconsistent concepts are more likely to occur if the set is general rather than specific. Consistent concepts are favored by the occurrence of identical elements based on meaning or group relationships, whereas inconsistent concepts are influenced by identical elements based on perceptual (and incidental) similarities.

Order of attainment. Some rather remarkable and significant behavior was observed in Heidbreder's experiments (**28** through **35**), which has been confirmed by Wenzel and Flurry (**74**). In these experiments, it was quite regularly found that concepts are formulated in a definite *order of dominance*, depending upon the nature of the material and the context. Under the conditions of these experiments the subjects first attained concepts of objects, then concepts of forms, and last, concepts of number. If, however, the stimuli are varied to make some critical features more obvious than others, there was likely to be a change in the order of dominance. Thus, in a series of twelve experiments involving card-sorting, varying conditions resulted in dominance of form, or color, or, in one case, quantity. When the form was represented by the whole card, which was also colored, and the concrete object was a small drawing on the card, the order of dominance was: color, spatial form, objects. Cards containing the forms outlined in different ways, with small objects varying in number and without color, were sorted in the order: spatial forms, objects, linear quantities, and number. On the whole, however, it was rather difficult to influence the order of dominance. In sixteen experiments where it was possible to attain concepts of objects,[12] these concepts were attained first eleven times, second four times, and last only once. Numerical quantities were attained last in ten experiments, second in two experiments, and first only once (in an experiment from which objects and forms were absent).

Concrete and abstract concepts. In experimenting with various sorting tests, Weigl (**64**) found that individuals characteristically displayed contrasting concrete and "categorical" (or more conceptual) approaches.[13] Using a variety of tests, Bolles (**2**) further analyzed these approaches, suggesting that concrete performance is primarily determined by sensory impressions whereas the more conceptual or abstract performance is a response to objects as representatives of a class or category. The "basis of pertinence" (or way in which objects are classified) was found to be of the following types:

1. *Identity*—items belong together by virtue of exact sensory equivalents.
2. *Partial identity*—similarly determined but resting upon an equivalence of only one sensory attribute.
3. *Co-functionality*—items belong together in some concrete situation.
4. *Categorical similarity*—items belong together because they represent a class, not because of similar attributes or contiguity of function.[14]

---

[12] Including the experiment of Wenzel and Flurry (**74**).

[13] It is interesting that English (**10**) reported substantially similar approaches.

[14] In this research, and others below, the primary emphasis was upon the comparative conceptual performance of aments and/or dements and normal persons.

Hanfmann and Kasanin (**20, 22**) developed a standardized procedure for testing conceptual performance with the Vigotsky sorting test. This test consists of a collection of small blocks, varying both in color and shape. On the bottom of each block is a nonsense syllable which serves as a clue to the proper classification of blocks into a double dichotomy (tall-large, tall-small, flat-large, and flat-small), the basis for which is not immediately obvious to the subject. After preliminary explanation of the purpose of the test the experimenter turns up one of the blocks, thus revealing its name, and asks the subject to select all those that belong with it. If the subject selects all those of the same color, for example, the experimenter turns up one of the wrongly classified blocks and asks the subject to try again, and so on until all the blocks have been correctly sorted. It is possible to score performance in terms of time taken, number of clues (turned-up blocks) required, type of solution, etc.

Through the use of this test, these investigators have distinguished three levels of performance, namely: a primitive, "concrete" performance, an intermediate performance showing some aspects of conceptual thinking but not others, and an advanced conceptual performance. For example, with regard to interpretation of the task, the subject at the primitive level does not relate the words to the properties of the blocks, with the result that the names of the blocks are disregarded and the task may be seen as solvable only by trial and error. At the intermediate level the subject does not grasp fully the nature of the required grouping; i.e., the task may be seen as a game with rules, or only one possibility of grouping is seen. At the conceptual level, the task is comprehended as involving classification. Different kinds of groups tend to be formed by individuals operating at the three levels. At the primitive level grouping may be merely random or in the form of constructions; at the intermediate level, the subject may assemble collections (such as one block of each color) or pseudo-classes (that is, all red blocks, where the basis can be shown to be "red," not color); and finally, at the conceptual level, true classes are formed.

Goldstein and Scheerer (**19**) have also used sorting tests to investigate conceptual behavior and have also found these contrasting approaches. They define the concrete attitude as realistic—that is, dependent upon immediate sensory impressions—whereas the abstract attitude is characterized by detachment from immediate reality and includes more than the "real" stimulus.

Although these investigators have been primarily concerned with pathological cases, this distinction between concrete and abstract performance may have considerable significance for normal, everyday behavior. Goldstein and Scheerer (**19**) point out that there are various degrees and gradations of these approaches. It is probable that "the normal individual combines both attitudes and is able to shift from one to the other at will according to the demands of the situation." More work with normal subjects is clearly needed in this area, since other experiments with different emphasis have strongly suggested that these attitudes are important aspects of concept formation and use. Hanfmann (21), in fact, has presented preliminary evidence based on the Vigotsky test for "personal patterns" of approach in tasks involving concepts. In the predominantly "thinking" or conceptual (abstract) approach, the subject develops hypotheses without paying much attention to the blocks, which are utilized to check the hypotheses; all of this may be accomplished without actually touching the blocks. On the other hand, in the predominantly "perceptual" (concrete) approach, the individual keeps in close touch with the material and gets his ideas from looking at or handling the blocks. A predominant approach of one kind or the other is characterized as a "concordant" pattern, whereas in a "discordant" pattern the two tendencies conflict with each other, leading to a less efficient performance.

Concluding Statement

In this review of concept formation we have been concerned to show that it will ultimately be necessary to take into account a wide variety of facts before an adequate formulation can be achieved. In general, concepts represent regulating (or selective) systems in the mental organization of the individual—systems which link separate sensory impressions, which depend upon past experience, and which are organized apart from direct sensory data. A given concept has both extensional and intensional meanings; it may have a consistent or inconsistent organization; it has important hierarchical relationships to other concepts. Furthermore, an adequate distinction must be made between the selective mechanism of the concept and the response selected—whether verbal (such as naming), manual (such as reaching for an object), or covert (such as an image).

So far as the available experimental data go, it appears that three major problems have been studied but without sufficiently clear differentiation between them. We have suggested that one problem is concerned with the development of the ability to conceptualize (to form and use concepts), another with the acquisition of a repertory of concepts (building up an organization of particular concepts), and still another with how any particular concept is achieved. The first two problems are basically genetic problems which must be approached primarily through tracing the learning process in the child; the third problem is, of course, important in child behavior, but it has been studied primarily at the adult level. Indeed, we have pointed out that it seems to be the only problem of concept formation which can be investigated at the adult level inasmuch as the adult has already developed an ability to conceptualize and already possesses so great a store of concepts that "new" ones are probably applications or reorganizations of previously learned ones. Nevertheless, it is still important to find out more than we now know about the characteristics of adult concepts, about how the repertory is used in a perceptual or learning or problem-solving situation.

Despite the magnitude of the problems, only a comparatively few methods have been used to investigate them. With children, the two methods have been an interview-questionnaire technique and a performance situation. Although the two genetic problems have not clearly been distinguished, as pointed out above, the latter method seems to have been applied mainly to Problem 1 (that is, the performance of the child is tested to see if he can conceptualize in various situations), and the former method to Problem 2 (that is, the child is asked questions in order to determine what concepts he has, how he uses them, etc.) With adults, the methodology has included introspection, learning situations, and problem-solving. The favorite procedure in the learning situation has been the anticipation method. Most experiments, however, have not included a preliminary learning period but have resembled problem-situations.

Bearing in mind the limitations imposed by these methods—for example, that the intensional and inconsistent aspects have been largely ignored—we can, nevertheless, note that a certain amount of progress has been made towards answering the three currently identifiable problems in the investigation of concept formation.

Problem 1, which concerns conceptualizing ability in general, has been the most inadequately investigated of the three. We do not really know for certain that there is an ability to conceptualize which might be distinguished from other perceptual and intellectual functions. Undoubtedly, children's concepts change with age, but whether or not this reflects development in a general function to form and use concepts has not been convincingly demonstrated. Several studies suggest that children can and do form concepts at a very early age and that the chief differences in this function between children and adults are of degree rather than kind. Thus, children have fewer concepts, perhaps because they have not yet had sufficient experience rather than because they do not yet have sufficient conceptualizing ability. Similarly, they cannot verbalize their concepts as well as can adults, but this does not necessarily mean that they do not have the concepts. An important finding in this connection, which warrants much more research, is that "concept scores" determined in various ways have a low correlation with intelligence test scores but a high correlation with training and experience.

With regard to Problem 2 (the acquisition of a repertory of concepts) much more is known. In fact, most investigations of children have dealt primarily with it. Here it has been found that concepts tend to change with increasing age, becoming more numerous, more complex, and more "logical." It has been tentatively ascertained that children acquire concepts in a definite order, although more systematic studies at very early ages are needed. For example, triangularity can be established as early as fifteen months, roundness by the age of three years, and size relations somewhat later. Concepts of time, social concepts, etc., are acquired later and undergo long development up into adulthood.

Finally, Problem 3 (the formation of a particular concept) has been dealt with more intensively than either of the other two. The great bulk of the research with adults has been concerned with discovering what happens when an individual evolves a particular concept or group of concepts out of a set of stimulus material. It has been found that performance of this kind is influenced by many conditions of presenting the stimuli, such as: (a) different ways of presenting the stimuli (order of presentation, varying the complexity; (b) contextual factors; (c) the inclusion of negative as well as positive instances; (d) set; (e) interference between similar series. In addition to these results bearing upon objective features of the concept-forming situation, what might be called "concept behavior" has been extensively observed. Among the characteristics noted are the tendency for subjects to evolve concepts gradually, the occasional attainment of concepts without awareness of the concept itself, the development of inconsistent as well as consistent concepts, the tendency for concepts to be learned in a definite order, and the occurrence of varying degrees of concrete and abstract approaches to, and use of, concepts.

Out of all these findings at least four significant points emerge for future investigators to consider.

1. The first is the distinction often mentioned above between child and adult concept-formation. The development of conceptualizing as a general function merits especial attention.
2. The second point is the surprising paucity of methods which have been employed to investigate the problems of concept-formation. The methodology, therefore, should be extended and elaborated.
3. The third point is the need to recognize more clearly than has been the case heretofore that concepts have internal and organizational characteristics as well as those objectively definable in terms of the external (e.g., verbal) behavior of the individual.
4. Finally, most investigators have concerned themselves only with very limited aspects of concept formation, so that research on a wider scale is needed along with conceptual integration of all the varied facts discussed in this article.

BIBLIOGRAPHY

1. Ames, L. B. The development of the sense of time in the young child. *J. genet. Psychol.*, 1946, **68**, 97–125.
2. Bolles, M. M. The basis of pertinence: A study of the test performance of aments, dements, and normal children of the same mental age. *Arch. Psychol., N. Y.*, 1937, No. 212.
3. Bruce, M. Animism vs. evolution of the concept "alive." *J. Psychol.*, 1941, **12**, 81–90.

4. Chant, S. N. F. An objective experiment on reasoning. *Amer. J. Psychol.*, 1933, **45**, 282–291.
5. Colby, M. G., & Robertson, J. B. Genetic studies in abstraction. *J. comp. Psychol.*, 1942, **33**, 385–401.
6. Dennis, W. Piaget's questions applied to Zuni and Navaho children. *Psychol. Bull.*, 1940, **37**, 520.
7. Deutsche, J. M. The development of children's concepts of causal relations. Univ. Minnesota, Inst. Child Welf. Monogr. Series. Minneapolis: Univ. Minnesota Press, 1937.
8. Dixon, J. C. Concept formation and emergence of contradictory relations. *J. exp. Psychol.*, 1949, **39**, 144–149.
9. English, H. B. *A student's dictionary of psychological terms.* (4th Ed.) New York: Harper, 1934.
10. English, H. B. An experimental study of certain initial phases of the process of abstraction. *Amer. J. Psychol.*, 1922, **33**, 305–350.
11. Ewart, P. H., & Lambert, J. F. The effect of verbal instructions upon the formation of a concept. *J. gen. Psychol.*, 1932, **6**, 400–413.
12. Fisher, S. C. The process of generalizing abstraction, and its product, the general concept. *Psychol. Monogr.*, 1916, No. 90.
13. Friedman, K. C. Time concepts of junior and senior high school pupils and of adults. *Sch. Rev.*, 1944, **52**, 233–238.
14. Friedman, K. C. Time concepts of elementary school children. *Elem. Sch. J.*, 1944, **44**, 337–342.
15. Gellermann, L. W. Form discrimination in chimpanzees and two-year-old children. I. Form (triangularity) per se. *J. genet. Psychol.* 1933, **42**, 3–27.
16. Gellermann, L. W. II. Form vs. background. *J. genet. Psychol.*, 1933, **42**, 28–50.
17. Gengerelli, J. A. Mutual interference in the evolution of concepts. *Amer. J. Psychol.*, 1927, **38**, 639–646.
18. Grigsby, O. J. An experimental study of the development of concepts of relationship in preschool children as evidenced by their expressive ability. *J. exp. Educ.*, 1932, **1**, 144–162.
19. Goldstein, K., & Scheerer, M. Abstract and concrete behavior; an experimental study with special tests. *Psychol. Monogr.*, 1941, No. 239.
20. Hanfmann, E., & Kasanin, J. A method for the study of concept formation. *J. Psychol.*, 1937, **3**, 521–540.
21. Hanfmann, E. A study of personal patterns in an intellectual performance. *Character & Pers.*, 1941, **9**, 315–325.
22. Hanfmann, E., & Kasanin, J. Conceptual thinking in schizophrenia. *Nerv. ment. Dis. Monogr. Ser.*, 1942, No. 67.
23. Hayakawa, S. I. *Language in action.* New York: Harcourt, Brace, 1941.
24. Hazlitt, V. Children's thinking. *Brit. J. Psychol.*, 1930, **20**, 354–361.
25. Hebb, D. O. *Organization of behavior.* New York: Wiley, 1949. (Espec. Ch. 6.)
26. Heidbreder, E. An experimental study of thinking. *Arch. Psychol.*, 1924, No. 73.
27. Heidbreder, E. Toward a dynamic psychology of cognition. *Psychol. Rev.*, 1945, **52**, 1–22.
28. Heidbreder, E. The attainment of concepts: I. Terminology and methodology. *J. gen. Psychol.*, 1946, **35**, 173–189.
29. Heidbreder, E. II. The problem. *J. gen. Psychol.*, 1946, **35**, 191–223.
30. Heidbreder, E. III. The process. *J. Psychol.*, 1947, **24**, 93–138.
31. Heidbreder, E., Bensley, M. L., & Ivy, M. IV. Regularities and levels. *J. Psychol.*, 1948, **25**, 299–329.
32. Heidbreder, E., & Overstreet, P.

V. Critical features and contexts. *J. Psychol.*, 1948, **26**, 45–69.

33. Heidbreder, E. VI. Exploratory experiments on conceptualization at perceptual levels. *J. Psychol.*, 1948, **26**, 193–216.
34. Heidbreder, E. VII. Conceptual achievements during card-sorting. *J. Psychol.*, 1949, **27**, 3–39.
35. Heidbreder, E. VIII. The conceptualization of verbally indicated instances. *J. Psychol.*, 1949, **27**, 263–309.
36 Hicks, J. A., & Stewart, F. D. The learning of abstract concepts of size. *Child Develpm.*, 1930, **1**, 195–203.
37. Hull, C. L. Quantitative aspects of the evolution of concepts. *Psychol. Monogr.*, 1920, No. 123.
38. Kuo, Z. Y. A behavioristic experiment on inductive inference. *J. exp. Psychol.*, 1923, **6**, 247–293.
39. Lacey, J. I., & Dallenbach, K. M. Acquisition by children of the cause-effect relationship. *Amer. J. Psychol.* 1939, **52**, 103–110.
40. Long, L. Conceptual relationships in children. The concept of roundness. *J. genet. Psychol.*, 1940, **57**, 289–315.
41. Long, L., & Welch, L. Reasoning ability in young children. *J. Psychol.*, 1941, **12**, 21–44.
42. Long, L., & Welch, L. Influence of levels of abstractness on reasoning ability. *J. Psychol.*, 1942, **13**, 41–59.
43. Moore, T. V. The process of abstraction. *Univ. Calif. Publ. Psychol.*, **1**, 1910, 73–197.
44. Munn, N. L., & Steining, B. R. The relative efficacy of form and background in a child's discrimination of visual patterns. *J. genet. Psychol.* 1931, **39**, 73–90.
45. Nagy, M. The child's theories concerning death. *J. genet. Psychol.* 1948, **73**, 3–27.
46. Oakes, M. E. Children's explanations of natural phenomena. Teachers College, Columbia Univ., *Contrib. to Educ.*, 1947, No. 926.
47. Ordan, H. *Social concepts and the child mind.* New York: King's Crown Press, 1945.
48. Peterson, G. M. An empirical study of the ability to generalize. *J. gen. Psychol.*, 1932, **6**, 90–114.
49. Piaget, J. *The child's conception of the world.* New York: Harcourt, Brace, 1929.
50. Piaget, J. *The child's conception of physical causality.* New York: Harcourt, Brace, 1930.
51. Piaget, J. Children's philosophies. In C. Murchison (Ed.), *A handbook of child psychology.* Worcester: Clark University Press, 1931.
52. Reed, H. B. Factors influencing the learning and retention of concepts. I. The influence of set. *J. exp. Psychol.*, 1946, **36**, 71–87.
53. Reed, H. B. II. The influence of length of series. III. The origin of concepts. *J. exp. Psychol.*, 1946, **36**, 166–179.
54. Reed, H. B. IV. The influence of the complexity of the stimuli. *J. exp. Psychol.*, 1946, **36**, 252–261.
55. Reichard, S., Schneider, M., & Rapaport, D. The development of concept formation in children. *Amer. J. Orthopsychiat.*, 1944, **14**, 156–162.
56. Russell, R. W., & Dennis, W. Studies in animism: I. A standardized procedure for the investigation of animism. *J. genet. Psychol.*, 1939, **55**, 389–400.
57. Russell, R. W. II. The development of animism. *J. genet. Psychol.*, 1940, **56**, 353–366.
58. Smoke, K. L. An objective study of concept formation. *Psychol. Monogr.*, 1932, No. 191.
59. Smoke, K. L. Negative instances in concept learning. *J. exp. Psychol.*, 1933, **16**, 583–588.

60. Smoke, K. L. The experimental approach to concept learning. *Psychol. Rev.*, 1935, **42**, 274–279.
61. Stevanović, B. P. An experimental study of the mental processes involved in judgment. *Brit. J. Psychol. Monogr. Suppl.*, 1927, No. 12.
62. Thorndike, E. L. *The elements of psychology*. New York: Seiler, 1905.
63. Thrum, M. E. The development of concepts of magnitude. *Child Develpm.*, 1935, **6**, 120–140.
64. Weigl, E. On the psychology of so-called processes of abstraction. (Trans. Margaret J. Rioch.) *J. abnorm. soc. Psychol.*, 1941, **36**, 3–33.
65. Welch, L. The development of size discrimination between the ages of 12 and 40 months. *J. genet. Psychol.*, 1939, **55**, 243–268.
66. Welch, L. The span of generalization below the two-year age level. *J. genet. Psychol.*, 1939, **55**, 269–297.
67. Welch, L. A preliminary study of the interaction of conflicting concepts of children between the ages of 3 and 5 years. *Psychol. Rec.*, 1938, **2**, 439–459.
68. Welch, L. A preliminary investigation of some aspects of the hierarchical development of concepts. *J. gen. Psychol.*, 1940, **22**, 359–378.
69. Welch, L. The development of discrimination of form and area. *J. Psychol.*, 1939, **7**, 37–54.
70. Welch, L., & Long, L. The higher structural phases of concept formation of children. *J. Psychol.*, 1940, **9**, 59–95.
71. Welch, L., & Long, L. A further investigation of the higher structural phases of concept formation. *J. Psychol.*, 1940, **10**, 211–220.
72. Welch, L. The genetic development of the associational structures of abstract thinking. *J. genet. Psychol.*, 1940, **56**, 175–206.
73. Welch, L., & Long, L. Comparison of the reasoning ability of two age groups. *J. genet. Psychol.*, 1943, **62**, 63–76.
74. Wenzel, B. M., & Flurry, C. The sequential order of concept attainment. *J. exp. Psychol.*, 1948, **38**, 547–557.
75. Wolff, W. *The personality of the preschool child*. New York: Grune and Stratton, 1947.
76. Woodworth, R. S. *Experimental psychology*. New York: Holt, 1938.

Received April 25, 1950.

METHODOLOGY IN PROPAGANDA RESEARCH[1]

T. R. VALLANCE
University of Massachusetts

Two major topics are taken up in this review. The first of these relates to the unsettled systematic status of propaganda, including the variety of ways in which propaganda phenomena have been conceptualized. In the course of developing this topic the framework for a methodological analysis of propaganda will be presented and illustrated. The second major topic concerns some of the more important efforts to study propaganda under experimental conditions and calls attention to some of the difficulties of such study.

THE UNSETTLED SYSTEMATIC STATUS OF PROPAGANDA

There is ample room for efforts to organize the facts and to theorize about propaganda, because no matter what one writes he can be fairly sure of not running counter to an experimentally verified and clearly organized body of fact and theory. A psychologist offering a neural perservation theory of "reminiscence" would promptly be shown to be at variance with established fact and theory. The thinker about propaganda, on the other hand, is almost free to conceptualize as he pleases without threat of serious professional consequences.

This section of the paper will illustrate several efforts to establish viewpoints concerning propaganda. In so doing it will illustrate the fact that a widely accepted systematic outlook in propaganda is lacking. Conceptualizations by different writers will be reviewed and several definitions of propaganda presented and discussed.

First, however, a note on the meaning of "conceptualization" as used in this paper will establish the framework of the following discussion. A *conceptualization of propaganda* is the specification of the independent, intervening (or intermediary), and dependent variables involved in propagandizing, together with statements or propositions showing how variables of the three classes relate to one another.

For purposes of the present analysis, the class of independent variables will include the doings of propagandists and the resulting propaganda stimuli which impinge on the propagandee.

The class of intermediary variables[2] includes (a) measures or indices of the social structuring of propagandees (*societal* intermediary variables) and (b) measures of the propagandees' attitudes, values, teleonomic trends, traits, needs, or other appropriate psychological constructs (*psychological* intermediary variables). Like intervening variables and hypothetical constructs in learning theory, variables in this class are thought of as bridging the gap between stimulus variables and response variables, explaining the nature of the relationship between them.

---

[1] The writer is indebted to Dr. Floyd H. Allport for certain trends of thought underlying this paper and for advice on certain aspects of the presentation.

[2] The term "intermediary variable" is suggested in place of the older "intervening variable" as used by Hull, Tolman, and other learning theorists. There are two reasons

The class of dependent variables here includes measures of the propagandees' responses to propaganda—change in attitude, memory for propaganda content, overt actions of specified kinds, etc.

It should be pointed out that the two sub-classes of intermediary variables, societal and psychological, while having the common function of explaining the connection between the stimuli presented by the propagandist and the responses of the propagandee, are not assumed to be on the same level of logical discourse. *Psychological* intermediary variables are in the nature of hypothetical constructs or intervening variables (cf. MacCorquodale and Meehl's discussion of these terms [30]) and as such are inferred from concrete data; they do not themselves generate the data. *Societal* intermediary variables, however, may be observed directly. That is, the data of the variable may describe, or otherwise refer directly to, objective facts within society. For example, one could compute an index based on knowledge of a person's socioeconomic status, religious affiliation, and other special group memberships, which might account for the differences between his and others' responses to common propaganda stimulus variables. Such an index, and the conditions it reflects, would not be "caused" or otherwise influenced by the propagandist (or the experimenter) and so would not be an independent variable as that term is used in the present analysis. (The index could, of course, be "manipulated" statistically once obtained, just as indices of psychological intermediary variables could.)

Likewise, the social structures mediating propaganda effects may also be considered as societal intermediary variables. For example, people may be urged to put pressure on their legislators in support of a propagandist's objective. The final "putting of pressure" depends on the existence and availability to the propagandees of the specialized social structures mediating communication (mails, etc.). Thus the un-

---

for this. First, data about certain variables in this class (e.g., indices of social structuring) are obtainable apart from the organism whose behavior is in question, whereas the intervening variables of the learning theorist are presumed to represent conditions or dispositions *within* the organism. Second, intermediary variables in the two major sub-classes are defined and measured by distinctly different procedures and are different in logical structure, as the following discussion brings out.

derstanding of this particular dependent variable ("number of urgings delivered to legislator X") hinges to some extent on a societal intermediary variable that could be indexed as "availability of communication channels to propagandees."

The choice of specific intermediary variables in either sub-class will, of course, be influenced by the particular dependent variable the experimenter wishes to examine, as well as by his particular theoretical preferences.[3]

The several conceptualizations discovered for discussion in this paper seem to fall into three fairly distinct categories. These categories may be briefly characterized before taking them up separately in detail:

1. *Linear conceptualizations* tend to suggest that propagandas operate as forces, each trying to push its recipients in some particular direction. In such conceptualizations emphasis is mainly on such independent variables as the content itself or the medium of its delivery, with relatively little attention being given to intermediary variables such as the social structuring of the propagandees and their psychological characteristics. Seldom, too, is effort given to specifying systematically the nature of the outcome, or dependent variable.
2. *Psychological conceptualizations* emphasize psychological intermediary variables such as the attitude and value systems of the propagandee and give little attention to the content of the propaganda or the propagandees' social structuring. Dependent variables are most commonly construed as psychological data: attitudes, opinions, beliefs, etc.
3. *Situational conceptualizations*[4] stress intermediary variables comprising the objective social structures in which propaganda must operate. Other independent variables and psychological intermediary variables are given less attention. Dependent variables are commonly construed as some form of social action as well as psychological data.
4. Occasionally a writer whose offerings are best placed in the situational category edges over into a fourth class of conceptualization which may be called *structural*. Here the effort is made to imagine all the possible variables that could be related to the effectiveness of a propaganda effort. The meaning of this as yet practically unoccupied category will be developed later.

Quite obviously, the discussions in the sources to be cited are not arranged so as to permit unambiguous identification of the variables the writers had in mind. But in organizing the available materials the present writer has tried to ferret out the others' meanings and to consider them in the framework just described, in order to compare them more easily and to evaluate them more fairly. Terms identifying the variables apparently had in mind by the various writers are bracketed within quotations.

---

[3] Relevant to our discussion of societal intermediary variables, an incidental topic in an earlier paper by Tolman (**41**) points somewhat in the same direction.

[4] The term "situational" is used in preference to "social," "societal," or "cultural" because of the intention to refer to the social situation (or social structuring) of the propagandee, rather than to the broader conditions characterizing the total culture in which the propaganda occurs.

Linear[4a] *Conceptualizations*

The clearest example of a linear view of propaganda is that of Moss (35), who says:

> Propaganda is an indirect type of appeal through suggestion [independent variable], which makes possible the management [dependent variable] of people. Propaganda may sell [dependent variable] a commodity, an idea, or a man—not on the basis of merit but through emotional and suggestive appeals [independent variable] which strike a sympathetic chord in the imagination [figure of speech?] of the public. . . . So powerful is political propaganda when skillfully manipulated [independent variable] that the mass of voters can be made to believe [dependent variable] one thing today, a quite different thing next week, and its opposite a month hence (pp. 424–425).

Here propaganda is a force, presumably taking the form of words, that may be turned on or off according to the desires of the person controlling it.

Another linear conception of propaganda is portrayed in *The Fine Art of Propaganda* (46), published under the auspices of the Institute for Propaganda Analysis. Here propaganda is defined as "opinion expressed for the purpose [independent variable] of influencing reaction [dependent variable] of individuals and groups," and the bulk of the book is devoted to illustrating in Father Coughlin's speeches "The Tricks of the Trade," the now well known seven "devices" that propagandists are said to apply in influencing the action of others. The "devices" are treated as tools that may be used at the whim of the propagandist, and it is pointed out that "Once we know these devices well enough to spot examples of their use, we have taken a great and long step toward freeing our minds from control by propagandists" (p. 24). Throughout the book no reference is made to what effect personal characteristics or social structuring of the victims of the devices may have on reactions to propaganda. The emphasis is entirely on the forces inherent in the "Tricks of the Trade."

Efforts at content analysis, exemplified in some of the work of Lazarsfeld and Merton (26) stem from an essentially linear conception of propaganda, though other works by these two writers fall well into other categories. Lazarsfeld and Merton define propaganda as "any and all sets of symbols [independent variable] which influence opinion, belief, or action [dependent variable] on issues regarded by the community as controversial" (**26**, p. 58). They then report studies in which definite features of propaganda—that is, features of the symbols—are linked with definite types of response. In their procedures, the symbols making up the propaganda are analyzed for particular themes, classified as favorable or unfavorable with respect to some propaganda objective; or the interrelationships of themes are studied. Such characteristics of propaganda are then correlated with "response analysis" data obtained through interview. Though a great emphasis on this type of approach to propaganda could lead to de-emphasis of social and psychological characteristics of the propagandees, Lazarsfeld and Merton recognize the limitations of this approach, and incidentally add a *situational* meaning to their conception of propaganda, in pointing out that "no propaganda [independent variable] can prevail [dependent variable] if it runs counter to events and forces [societal-psychological (?) intermediary variable] underlying those events. . . . Propaganda is no substitute for social policy and social action . . . " (p. 79).

[4a] The concept of the "linear" approach in psychological methodology was formulated by Floyd H. Allport (unpublished work) as a means of distinguishing the more customary logic from a "structural approach" upon which he is now working. The application of the concept, made here to propaganda studies, including the dependent-independent variable logic as an earmark of linearity, was developed by the present writer.

Another point of view on propaganda illustrating mixture of two categories, this time linear and psychological, is that of Lindesmith and Strauss (**29**). These writers say that propagandizing "refers to the attempted [independent variable] conversion [dependent variable] of masses of men and women to a point of view. Successful propaganda [independent variable] ends in mass results [dependent variable]: it causes numbers of people to act together as a group, or at best to act in similar ways. It may also be used to undermine the morale of opponents" (pp. 499–500). They then go on to list "techniques" the propagandist may use to get his results. However, in the discussion of the limitations of propaganda, their conceptualization appears to fall on the borderline between the linear and psychological categories: "The propagandist must play [independent variable] on the response system of people [psychological intermediary variable] and sometimes he will find his human subjects very recalcitrant [dependent variable]" (p. 503); this idea is then elaborated somewhat. Still less recognition is given to situational factors as a limitation on the propagandist when Lindesmith and Strauss point out that he may run against the propagandees' conceptions of themselves as fathers, churchgoers, and loyal citizens if he attacks basic institutions such as the family, church, or government. But this admixture is still largely psychological in that it deals with various concepts of self that may arise from institutional role-playing.

Strengths and weaknesses peculiar to linear conceptualizations. Here the major strong point seems to be the development of methods of analyzing propaganda content and of ways of recognizing how various utterances in public communication channels are slanted toward certain viewpoints. Curiously, though, this advantage grows out of the principal weakness of linear approaches, namely the overemphasis of one kind of independent variable—the verbal output of propagandists. More work could well be done with content analysis, but it might be more fruitful to give additional effort to identifying the particular psychological intermediary variables to which the various categories of content might be relevant. Efforts in this direction could well take "response analysis" (**26**) procedures as a point of departure.

There are other weaknesses in linear conceptualizations common to other conceptualizations which, to avoid repetition, are taken up after the others are considered (see pp. 43–45).

Psychological Conceptualizations

Psychological conceptualizations of propaganda are those more clearly emphasizing the attitudes, values, and other personal characteristics of propagandees (both as psychological intermediary variables and as dependent variables). The most clear-cut example of a purely psychological approach to propaganda is that of Krech and Crutchfield (**23**). In its very nature phenomenological, the psychology of Krech and Crutchfield cannot give emphasis to social structurings or to the symbol-content of propaganda apart from the way those things relate to the individual's perceptions as determined by his needs and values. Response to any stimulus, propaganda content included, depends solely on the cognitive reorganization of the "psychological field" of the individual at the time of stimulation. The psychological field, they say, is "a cross section, at a given moment, of the psychological existence of the individual. It includes only what is present at the moment, not what was present previously or what may be present in the future" (p. 33). Elsewhere, Krech and Crutchfield identify the psychological field as "a state of neuropsychic processes that are reflected in the form of the *experienced* world of the individual and in the form of patterns of *neural processes in the brain* of the individual. The psychological field is not identical with the experiences of the individual for there are important constituents of the neuropsychic process that are unconscious, i.e., do not result in immediate experiences by the individual" (p. 37). In their treatment of propaganda, then, these writers look upon words as objects in the real world, objects as real as the things for which they stand, and the propagandist thus appears as a man who can alter other people's worlds by introducing words into their environments just as he might by introducing people, machines, or furniture. "Propaganda analysis" for them would be mainly "concerned with the current beliefs, needs, percepts, attitudes of the listener" (p. 329). In a very subjective sense, then, their approach is purely psychological, dealing with the private experiences of the person. No effort is made to relate propaganda efforts and effects to objective collective relationships of the individual, for these things may or may not be important, depending on the perception of them by the person. Krech and Crutchfield are not clear about how the psychological field of a person at a given instant is to be ascertained for purposes of prediction, but that is not a problem for this paper.

Another viewpoint predominantly psychological, although having noticeable strains of *linearity*, is that of Doob (15). Doob defines propaganda as "the attempt [independent variable] to affect the personalities [psychological intermediary variable?] and to control the behavior of individuals toward ends [dependent variable] considered unscientific or of doubtful value [psychological intermediary variable] in a society [societal intermediary variable] at a particular time" (p. 240). The last seven words of the definition convey the only implication in Doob's treatment about the importance of social organization in the study of propaganda. He develops this phrase only in an effort to distinguish, as many writers do, education from propaganda. But beyond this differentiation, Doob's treatment of propaganda is almost strictly psychological, as suggested by the titles of his chapters 14, 15 and 16: "The Perception of Propaganda," "Personality and Propaganda," "The Learning of Propaganda." Chapter 13, "The Content of Propaganda," suggests linearity of approach, for here we find methods for determining and adjusting the relative proportion of various appeals and themes from which the unit of propaganda derives its strength.

To illustrate further Doob's psychological viewpoint we consider some of the details of the chapters just mentioned. In chapter 14, Doob speaks of auxiliary responses, roughly equivalent to attitudes, whose function is to produce the perception at the basis of the learning of propaganda. The propagandee's action results from "propaganda he has heard and from other attitudes within him. The responses facilitating perception have been merely auxiliary though absolutely necessary for the final behavior" (p. 321). Also in this chapter Doob takes up such standard topics in perception as stimulus intensity, stimulus context, stimulus simplification, perceptual repetition, and perceptual variation, and illustrates how these factors operate in various forms of propaganda from advertising to psychological warfare. In chapter 15 Doob emphasizes personality characteristics that the propagandist finds related to the propaganda objectives (that is, a "personality response" which is "instrumental in achieving a goal response or which mediates such a response. . . . In the case of propaganda the goal response to which a personality response is related or unrelated is the propaganda objective"). Other related material comes in chapter 16 where Doob mentions three reasons why personality responses aroused by propaganda may not result in the recipient's learning a new attitude. These are (1) their drive strength may be weak, (2) the responses may be unrelated to the propaganda goal, and (3) other personality responses previously aroused in connection with the propaganda goal may have greater drive strength.

Further research may show that an idea suggested in the material just cited represents a real advance in propaganda theory, viz., the relationship of personality characteristics to *action in bringing about* the propaganda objective. The role of personality in this context has received little attention, more study having been devoted to characteristics that may be associated with the propagandee's *becoming convinced of the desirability of the objective itself*.

A. M. Lee (**27**) mentions, among others (see below, pp. 41–42), a psychological approach which expresses the idea that social struggles and their propagandas attract personality types that can find needed expression in the resulting appeals, promises, and actions, and that such personalities once attracted have strengths and limitations that are felt throughout the entire conflict. Though Lee advances no evidence for this view, it is an interesting suggestion, similar to the one mentioned in the preceding paragraph, and one that should be specifically tested in an experiment.

Another largely psychological conception of propaganda, though this time with situational leanings, is given in the recent compendium of Smith, Lasswell, and Casey (**40**). Their statement that the "content and control of propaganda are directly related to social structure" (p. 2) suggests a situational type of orientation. But their definition of social structure as "the basic values of a given society and the pattern followed in distributing them" (p. 2)—together with the observation that "limits are put on the public media by the attitudinal standards of the community as a whole, which include morality, loyalties, expectations" (p. 2)—reveal a strongly psychological orientation. Although some recognition is given to collective involvements of the members of the public, these involvements are seen as important only insofar as they produce attitudes and values in people. The relative emphasis of collective vs. individual determinants of response to propaganda leads us to classify the approach as psychological.

Bartlett's (**6**) view is likewise psychologically oriented. He defines propaganda as "an attempt [independent variable] to influence opinion and conduct—in such a manner [independent variable] that the persons who adopt [dependent variable] the opinions and behavior indicated do so without themselves making any definite search [psychological intermediary variable] for reasons" (pp. 5–6). The essential process in propaganda is, for Bartlett, suggestion—largely prestige suggestion—which releases certain tendencies [psychological intermediary variable] already set to go. Bartlett does not discuss social or group allegiances of people as importantly related to responses to propaganda, and he attempts a psychological distinction between propaganda and education. The latter he sees also as attempts to influence thinking and conduct, but in education the individual is encouraged to seek to understand for himself why he does what he does.

Another psychological approach is that of Bird (**8**), who defines propaganda as "planned and systematically applied suggestion, expressed largely in symbols [independent variable], for the purpose of controlling attitudes and securing a predetermined mode of conduct [dependent variable]" (p. 310). Bird states further that "Propaganda is above all else addressed to and is successful through feelings and emotions [psychological intermediary variable]" (p. 313). Other predominantly psychological views are those of Kimball Young (**45**), A. L. Edwards (**19**), and E. L. Bernays (**7**).

Strength and weaknesses peculiar to psychological conceptualizations. The advantage in psychological approaches to propaganda lies obviously in the search for important psychological intermediary variables to tie together propaganda stimulus and propagandee response. After all, it may be argued, it is an organism whose response is the center of interest in propaganda as in other fields of behavior, and the more we can know about the psychological attributes of that organism the higher will be the accuracy with which we predict its behavior. This, essentially, seems to be the view of phenomenologists, who, in theory, frankly discount the importance of conditions not a part of the organism's psychological field or life-space.

As noted above, Doob has suggested the utility of studying personality characteristics that might predispose the propagandee *to do things toward* a propaganda objective in addition to those predisposing to *believing in the merits of* a propaganda objective; this is a type of psychological intermediary variable not suggested by other writers. Further research may show this to be a positive value in the study of psychological variables in propaganda.

The concentration upon psychological intermediary variables in psychological conceptualizations means that societal intermediary variables tend to be overlooked. While it may be taken as a sort of programmatic assumption that most psychological characteristics arise from experiences in social structures, social psychology still lacks specific and detailed knowledge of the relationship between the objective facts of social structures and various psychological phenomena. This means that we cannot justify high confidence in predictions of propaganda effects (and other behaviors) in the absence of measures and indices of relevant social structures made apart from, and considered in addition to, psychological variables. A second weakness in typical psychological conceptualizations (Doob's being an exception) is the relative lack of concern with independent variables per se; although in the over-all study of propaganda these variables are strongly emphasized, the relationship between them and psychological intermediary variables deserves more study.

Situational Conceptualizations

Situational conceptualizations of propaganda give relatively greater consideration than the two preceding types of treatment to the importance of collective organizations in which the propaganda recipients are involved. It should be re-emphasized that the categories around which the present discussion is organized are not mutually exclusive but differ from one another mainly in degree of emphasis on the various classes of variables in terms of which propaganda phenomena may be analyzed and upon which the prediction of the propagandee's response may be based. The present category includes views with considerable psychological emphasis, but they appear to stress social structures relatively more than those given earlier.

Illustrative of the situational viewpoint is that of LaPiere and Farnsworth (**24**), who point out that the direct effects of propaganda have often been exaggerated. They advance the view that it is "to the social conditions [societal intermediary variable] which make people susceptible to conversion [dependent variable], rather than to the propaganda itself [independent variable], that we must first look for an explanation of political or other leadership which has used propaganda to secure leadership status" (p. 447). While recognizing the importance of the psychology of the propagandees ("If they are reasonably content with what they have, propaganda will not be very effective," p. 447) they also recognize, by implication, that the peculiar individual "psychologies" arise from the social structurings of the time and place.

Another situational approach is followed by A. M. Lee (**27**). He first defines propaganda as "the use of words, symbols, ideas, events, and personalities with the intention [independent variable] of forwarding or attacking [dependent variable] an interest, cause, project, institution, or person [societal intermediary or dependent?] in the eyes and minds of a public [dependent or societal intermediary?]" (p. 127). After this rather loose definition, Lee describes a number of approaches to propaganda, the first of which is the "societal approach." Here he makes the point that propaganda arises and participates in social tensions and struggles and that environmental conditions set limits to societal change which in turn control the effectiveness that propaganda may have. Lee says that the investigator or propagandist should study (a) the nature of underlying social tensions and struggles and (b) who is competing with whom and why.

Other situational approaches are the following: Parsons (**37**) defines propaganda as one kind of attempt [independent variable] to influence attitudes and hence the actions of people [dependent variable] by linguistic stimuli [independent variable], and states that propaganda [independent variable] if it is to be successful [dependent variable] must be developed in relation to both the individual [psychological intermediary variable?] and the social institutions [societal intermediary variable] in which it is to work. R. H. Turner (**42**) also says that propaganda effectiveness [dependent variable] depends on techniques [independent variable] and the receptiveness of the social situation [societal intermediary variable] and interaction with the particular social situation [societal intermediary variable].

The clearest attempt reported thus far actually to study propaganda in a real life social situation is that of Lazarsfeld, Berelson, and Gaudet (**25**). Although this report contains scarcely any direct discussion of approaches to propaganda per se, the methods and data that are presented reveal the situational approach in its clearest form. This study of "how the voter makes up his mind in a presidential campaign" describes, in terms as objective as the available methods permit, the institutional allegiances, economic status, and family loyalties which were related to voting behavior and the effects of the propaganda campaign upon such behavior. In other words, the study carefully analyzed the social situations in which a propaganda war was being waged, and sought, with considerable success, to identify those components of the situation related to several propaganda effects. The methods and some of the results of the study are reported in more detail in the next section.

Strengths and weaknesses peculiar to situational conceptualizations. To this writer there seem to be two main strengths in situational conceptualizations. The first of these is in the concentration on social structuring (or societal intermediary variables), which is not only the scene of propaganda activities but constitutes some of the major conditions which produce the frustrations and anticipations of frustrations giving rise to propaganda in the first place. It seems likely that no full understanding of propaganda can be had without recourse to knowledge of social structures likely to be influenced by the attainment of propaganda objectives and of the roles of propagandees in those structures. This leads to the second apparent strength in situational conceptualizations: a tendency toward the recognition that the dependent variable in a propaganda campaign is eventually to be reflected in some realignment of social forces or objective change in some aspect (however small or great) of the social structure.[5] This being so, dependent variables are to be predicted, in part at least, from indices based on what are here called societal intermediary variables. The use by Lazarsfeld, Berelson, and Gaudet (25) of an "Index of Political Predisposition" is an example of this.

Among weaknesses in situational conceptualizations may be noted a lack of concern with independent variables—content and the doings of the propagandist—and relative lack of systematic concern with psychological intermediary variables (understandable in consideration of the training and interests of most writers on propaganda whose products would be classed as situational). Also, much work remains to be done on how to assess for the propagandee his particular involvements or roles in various structures of society in order to relate clearly these involvements and roles to his response to propaganda (a version of one of the perennial problems of social psychology).

Summary of Methodological Weaknesses Common to Conceptualizations Considered Thus Far

The following paragraphs have a generally critical tone, so let it first be emphasized that the conceptualizations reviewed above when taken together contain much that is good and useful, though it is not yet available in systematic language and organization. This section emphasizes common methodological weaknesses in the hope that past and present shortcomings may be avoided in a more positive contribution.

1. First, and most general, definitions and discussions of propaganda tend to be quite loose and "global," often failing to recognize the need to distinguish the three major classes of variables which must be specified before research can proceed systematically. For example, when Lee (27) defines propaganda as "the use of words, symbols, ideas, events, and personalities with the intention of forwarding or attacking an interest, cause, project, institution, or person in the eyes and minds of a public" it is difficult to know what variables are meant to be indicated by these phrases. Should the independent variable be "use of" or "use of words . . . and personalities," or should "intention" be included? Perhaps "intention" would be a part of a societal intermediary variable, since it implies (with " . . . of forwarding or attacking a cause. . . . ") a context of social conflict. Perhaps "forwarding or attacking . . . " should be thought of as dependent variables since they may be outcomes of "use of. . . . " At any rate, unless a definition—accompanied by appropriate clarifying statements—really *defines* and *delimits*, it is of little use as a starting point for scientific research and understanding.

2. The treatment of independent variables (assuming they are clearly identified) is a more specific trouble area. As mentioned, content analysis is a well developed aspect of propaganda research. But other independent variable

---

[5] A comparison of definitions in this and the section on psychological conceptualizations will show greater emphasis in the latter on "belief, attitude, and opinion," whereas definitions in the present section more clearly imply overt action of one kind or another.

attributes should be explored more carefully: source, frequency, mode and context of content presentation; saturation (percentage of a given communication medium occupied by a propaganda); clarity, and the directness-indirectness dimension; specificity of action urged. All of these terms (and probably others could be found) represent independent variables which could be systematically experimented with for years provided one had a good conceptual system to work with.

3. The treatment of intermediary variables is a third trouble area. Sociologists and psychologists now have available a number of indices (see Appendix to **25** for several interesting examples) and tests which may be used to develop intermediary variables of both the societal and psychological types for use in propaganda and other social psychological experiments. Most propaganda experiments at all concerned with intermediary variables have used only one of the two subtypes discussed in this paper (cf. **20, 33, 39**). The studies by Lazarsfeld, Berelson, and Gaudet (**25**) and Merton (**34**), and an as yet unpublished experiment by the writer, have made some effort to understand propaganda campaign effects through the use of both kinds of intermediary variables. So long as both types have predictive value independently of one another, efforts should be made to integrate the two and to explore their interrelations more fully.

The phenomenological approach (cf. **23**), incidentally, seems to this writer to contain a risk of certain methodological confusions of psychological intermediary with dependent variables. If one emphasizes the importance of how-the-propagandee-perceives-the-propaganda as a determinant of his reaction to it, he must not overlook the fact that perceptions are known to the outer world only by some response. Strictly speaking, this informing response (informing about the perception of the propaganda—not about the perception of other, possibly predictively related conditions) is a dependent variable and cannot legitimately be used to infer some previous construct from which it is itself to be predicted or explained ex post facto. The predictor must be observed or inferred independently of the predicted.

4. The treatment of dependent variables is a fourth trouble area needing careful systematization. A review of the several definitions of propaganda given on previous pages will show that such terms as *opinion*, *sell*, *attitude*, *action*, *reaction*, *conversion*, and *undermine* are used to point to outcomes of propagandizing. Little effort is given to specifying how such terms may be made operational (i.e., their referents measured)—viz., how a "reaction" is to be identified, how a public is to be recognized as "sold" on some idea, or just what constitutes a "conversion." Lazarsfeld, Berelson, and Gaudet (**25**) may again be cited as having been wide awake to the various outcomes which a propaganda campaign may have.

In the last analysis, the success of a propaganda campaign is to be revealed in some form of social action, and the researcher must be ready to specify how the particular dependent variables (e.g., attitudes, beliefs, etc.) he is studying in his particular situation are related to action—this is a region for some very profitable basic research. Doob's (**15**, chap. 4) analysis of auxiliary response-perception-learning-action appears definitely as a step in the right methodological direction. The whole problem of relating attitudes, beliefs, needs, opinions, etc. to overt action is of course a central one in psychology; its solution is being actively sought by researchers of several theoretical persuasions and would materially advance our understanding not only of propaganda but of many other problems as well.

5. The main evidence that the study of propaganda is still at a rather empirical or common sense level is in the absence of statements or propositions showing how variables of the different classes are related to one another. Linear conceptualizations are highly empirical, sometimes overlooking altogether not only the problem of relationships between variables but the very existence of intermediary variables. What is needed, for example, is a clear notion of how and why a given mass of propaganda stimuli (independent variables) are perceived in the first place: (a) what personality characteristics, values, attitudes, drives, etc. (psychological intermediary variables) contribute to perception or failure of perception, (b) how a given perception gives rise to a measurable attitudinal response (dependent variable), (c) how attitude (or belief or opinion) is related to other behaviors (dependent variable), and (d) how involvement (membership, role-playing) in a given set of groups and institutions (societal intermediary variables) is translated into psychological variables and thence into response. Psychology is far from barren of resources, both technical and theoretical, for handling such problems (e.g., see **15, 23**), and it is not the present writer's intent to imply that no work has yet been seriously undertaken toward solving them. Rather it is his hope that more actual experimentation in genuine propaganda situations—in addition to pilot studies of a laboratory sort (such as **28**)—might be undertaken to examine the significance of different variables and the relationships among them.

The Need for a Structural Conceptualization

In general terms, we mean by a "structural conceptualization of propaganda" the following: a set of definitions and propositions which endeavors to give adequate consideration to variables of all classes, and to encompass advantages already noted in the linear, psychological, and situational conceptualizations related to the persuading of large numbers of people toward some objective. This view stresses especially the relationships among those variables—how they are *structured together* to produce a propaganda campaign.

A rather unsystematic and yet broad view of propaganda is presented in a recent essay by Flowerman (**18**) which comes fairly close to the notion of a structural approach. Although Flowerman's article deals with propaganda in the "war against bigotry," what he has to say may well be considered in application to propaganda in general. Flowerman takes up three general topics, each with a number of subtopics: (1) conditions necessary for mass propaganda to be effective, (2) content of propaganda, (3) effectiveness of propaganda.

1. Under the first topic he lists *control* by the propagandist of the channels of mass communication, *saturation* (referring to the proportion of the total mass of communication made up by the particular propaganda under considera-

tion), *focussing of attention* (influenced by mode, source, context and frequency of symbol presentation), *comprehension* of the propaganda by the receivers, and *conformity* of the propaganda to group standards (i.e., standards of groups with which propagandee identifies). These factors clearly include conditions not made at all explicit in previously cited treatments of propaganda; they include factors both within the individual (e.g., comprehension), outside him (e.g., saturation), and his particular social relationships (e.g., conformity of propaganda to group standards).

2. Flowerman's discussion of propaganda content is brief and not particularly enlightening as compared with what other writers, notably Doob (15) and Lazarsfeld and Merton (26), have given us. He does point out a need for clear-cut, unambiguous presentation of the propaganda objectives and cites the experiment of Cooper and Jahoda (12) and an unpublished study by Berelson as evidence that people unsympathetic with the objective of the propaganda may misinterpret or fail completely to understand the propaganda if it is vague or ambiguous.

3. Flowerman's discussion of ways of evaluating the effectiveness of propaganda includes criteria implied in many definitions of propaganda: voting, writing letters to legislators, persuading others, writing in for more propaganda. He also mentions another measure implied by many writers and used in many experiments related to propaganda: scales for the measurement of attitudes and information. But in addition Flowerman suggests immediate tests of recall of the propaganda, delayed recall tests, and repetition of the propaganda to others.

Flowerman's statement, while containing no effort at systematic analysis into variables of different classes, makes explicit a number of things clearly relevant to a going propaganda program which are mentioned only piecemeal here and there by other writers. His treatment includes many items which should go into what we are calling a *structural* conceptualization, whose major requirements may be stated as follows:

1. A structural conceptualization of propaganda would consider the effects which the attainment of the propaganda objective would have on groups and institutions important to the propagandee; that is, *it would relate dependent and societal intermediary variables.*

2. It would study personality characteristics of the propagandee both as they relate to (a) acceptance of the propagandist's arguments and (b) taking available actions toward bringing about the propaganda objective; that is, *it would relate dependent and psychological intermediary variables.*

3. A structural approach would also include a study of propaganda content, as the appeals used relate to personality characteristics, abilities, and interests of the propagandee; and it would consider the manner and medium of the delivery of the propaganda as related to personality characteristics of the propagandee; that is, *it would relate independent and psychological intermediary variables.*

4. Finally, and most crucially, it would endeavor to provide a means for studying how such factors are empirically structured together to produce

propaganda effects; that is, *it would have to contain propositions showing clearly the relationships among variables of the three classes mentioned.*

With such a framework set up, and adequate instruments for the measurement of appropriate variables, it would become possible to conduct a long series of experiments to explore systematically a large number of conditions of suspected importance in the understanding of propaganda. Quite obviously such a program would require theoretical orientation in terms of broader social and psychological theory, and a long series of pilot experiments aimed at developing measuring devices, before serious systematic exploration could begin.[6]

EXPERIMENTAL INVESTIGATIONS OF PROPAGANDA-PERSUASION PROCESSES

The purpose of this section is to illustrate techniques of research in propaganda and to point out difficulties in this field of research; the section does not present an exhaustive survey of the literature. The section contains three parts: a review of studies of attitude change, a review of studies of action, and a brief summary of problems and shortcomings encountered in research to date.

Studies of Attitude Change

The experimental modification of attitudes typically involves a control group and one or more experimental groups, all equated at the beginning of the experiment with respect to some variable—most frequently the dependent variable (i.e., attitude). Control of differential effects of extraneous variables is most frequently achieved by randomly assigning subjects to groups in the experiment. The experimental groups are then subjected to the predetermined conditions, all groups are compared on the dependent variable some time later, and any differences are attributed then to differences in experimental treatment. Occasionally control groups are not used if the dependent variable is judged to be normally quite stable (**38**). Independent variables applied as experimental conditions have included emotional vs. rational speeches (**21**), majority vs. expert opinions (**31**), favorable vs. unfavorable reading matter (**5**), speech vs. radio address vs. reading (**44**), and various types of school instruction (**22, 16**).

---

[6] Although there is nothing in print today which begins to meet the requirements of an integrated structural conceptualization, it is the writer's understanding that the current work of Floyd H. Allport and his students is pointed in this direction. (This work is for the most part unpublished as yet, but see **2** for an early and now somewhat out-of-date statement.) A set of hypotheses and a rationale for a general structural approach to the study of behavior are being prepared by Allport for publication.

Reviewing the available studies of relationships between propaganda and attitude changes, one cannot help being struck by the surperficial nature of most of the variables examined. This may be due in part to the apparent (though misleading) ease with which experiments of this kind may be set up and conducted, but probably more to a lack of a systematic theory of attitude. The available studies are almost entirely concerned with what we have previously described as a linear approach, with only occasional effort being made to relate experimentally produced changes to anything other than the manipulations of the propagandist. No report seen by this writer deals with a propaganda experiment using attitudes as dependent variables in which the attributes of the psychological, situational, or structural approaches as described earlier were given serious consideration. The fact of "unreality" of the experiments which Murphy, Murphy, and Newcomb (**35**, chap. XIII) lament is not nearly so important as this lack of study of inter- and intra-individual factors that might be related more importantly to attitude change than most variables made up of symbolic content or other non-coercive social experiences to which the subject in the experiment might be exposed.

If one were to push further Murphy, Murphy, and Newcomb's reservation about unreality of attitude-change experiments, he would have to raise the question of the relationship between attitudes and action. While these experiments are relevant to propaganda definitions stressing attitude and opinion, they offer nothing conclusive about overt action as a result of propaganda. McNemar (**32**) raises the question of validity of attitude measurement, suggesting action as a criterion of validity (p. 296). The typical attitude-change experiment does not raise the question of the attitude-action relationship but tacitly assumes one to exist, attitude being the presumed determining factor. Until more conclusive evidence than we now have is at hand concerning a *causal* relationship between attitude and action we must remain very skeptical of the significance of even the most sophisticated experiments in which attitude measurement comprises the dependent variable.

A study by F. H. Allport and Simpson (**2**) of Axis broadcasts to the United States during the war makes use not of attitude but of degree of belief of assertions in the propaganda, such as (a) that the U.S. was on the verge of defeat and (b) that our allies were aiming to exploit and betray us. The use of a measure of belief as a dependent variable is an interesting departure from previous procedure, but it too needs "validation" against action measures.

An as yet unpublished experiment falling on the borderline between those using measures of attitude or belief on the one hand and those using measures of action on the other is that of Franklin (**17**). In this experiment, conducted during the recent war, three groups of college student subjects were given three sessions comprising about ninety minutes of recorded propaganda. One group received United Nations appeals, chosen from those used in the Deren experiment reported on pages 51–53 below; a second group received Axis appeals, selected and edited from German and Italian broadcasts in March and April of 1942; and a third group served as a control group, receiving no appeals. Prior to the propaganda each subject executed a Civilian Participation Inventory comprised of seventy-one items on which he could indicate his participation (or non-participation) in civilian war activities. The subject could indicate his participation (as of the past one-month period), his *intention* to participate in the future, or both, on any item of the Inventory. Between the first execution of the Inventory and the propaganda, the subject copied in longhand (as a part of a concurrent experiment not reported in detail by Franklin) letters to be distributed to other students urging them to discuss and work in the war effort and to urge still others to do so. Following the three propaganda sessions was another letter-writing session after which the Civilian Participation Inventory was executed for the second time.

We need not detail here the kinds of indices obtained from the scoring of this inventory. The results are of interest, however, and are congruent with those reported by Deren (**13**, see pp. 51–53 below). All three groups increased their *participation* scores on the Civilian Participation Inventory, with differences significant at least at the 5 per cent level of confidence. The gain for the group receiving the United Nations appeals was significant at the 1 per cent level. There were *no significant differences*, however, between the *gains* of the groups. Therefore, the group gains could not have been due to any differences in propaganda content, especially since the group receiving no appeals increased in mean participation score as much as the others. There thus seemed to be other influential factors at work.

When *intention* score differences are considered, much the same general trend is apparent, though the gains were not so high as on the participation measures. The United Nations appeals appeared to be slightly, though not with full statistical significance, more effective than the pro-Axis appeals. The latter were a little more effective than no appeals at all.

Franklin interprets these results as providing "no evidence for the usual naïve belief in the absolute energizing effects of propaganda" (p.

175). She interprets them, rather convincingly, as being due to the effects of what is called "structure." It is noted that *all three* groups (a) executed the Inventory, (b) copied letters aimed at arousing others to war activities, and (c) were members of a nation at war, *before* the propaganda was received. These three things are seen as clarifying the fact that the subject was in a situation structured toward certain outcomes ("victory"). The appeals themselves were insufficient to increase (by comparison with *no* appeals) the amount of one's participation, or intention to participate, in civilian war activities. Nor did arguments directed toward participation in the war effort produce significantly more effect, or any different *direction* of effect, than the arguments designed by the enemy to lessen the subject's participation in the war effort.

Franklin then hypothesizes that propaganda "functions only as a servant of 'structure' [some form of societal intermediary variable], is dependent upon it, and is meaningful only in terms of it. It is at the mercy of 'structure' and thus can be completely neutralized by it" (p. 179). The last sentence in the quotation evidently refers to the fact that pro-Axis appeals also were followed by rises in participation scores.

The participation and intention measures as used by Franklin are interesting departures from previous usages in a realistic direction, namely, toward bringing the measure of the dependent variable closer to the reality demanded by successful propaganda: action.

Studies of Action

Turning now to experimental studies of propaganda in which some form of action is taken as the dependent variable, we find only four: those of Hartmann (20), Deren (13), Lazarsfeld, Berelson, and Gaudet (25) and Merton (34). The small number of such studies testifies to the practical and methodological difficulties inherent in them. The amount of time, effort, and money required to discover and measure the significant variables, not to mention the cost of the propaganda itself, is prohibitive. And the proper control of variables once they are discovered is indeed a tricky methodological problem.

Hartmann's (20) experiment has been reviewed extensively in other places (cf. 7, 34) and so will be treated but briefly here. In the 1935 Pennsylvania elections Hartmann himself was a candidate for office in Allentown, running on the Socialist ticket. There was agreement within the party that the Socialists were quite unlikely to win any offices, making possible the further agreement that nothing would be lost by using Socialist election propaganda in an experimental test of "emotional" vs. "rational" appeals. The experimental groups were the population of Wards 2, 9, and 10, which received "emotional" appeals, and of Wards 3, 4, and 7, which received the "rational" appeals. The people in the remaining 12 of the 19 wards served as controls and received no Socialist propaganda whatever. The groups of wards were matched as well as possible with respect to population density, socio-economic status, previous voting habits, size, and assessed real estate values. The appropriate leaflets were handed to the "subjects" by party workers on calls at residences (or left under doors of those not at home). The criterion of effectiveness was, of course, voting behavior. Results are presented in terms of relative gains in Socialist vote in the ward groups over the Socialist vote in the election of the preceding year: the over-all vote increased by 16.69 per cent, with the Republicans increasing 19.79 per cent, the Democrats 13.10 per cent, and the total Socialist vote 30.86 per cent. Experimental group differences are seen in the fact that the Socialist increase was made up of a 50 per cent increase in the emotional wards, 35.42 per cent in the rational wards, and 24.05 per cent in the control wards. Interview follow-ups after the election tended, in terms of recall of the leaflet content, to confirm the superiority of the emotional appeal.

Hartmann himself calls attention to some of the weaknesses of this study, principally the failure to match the "emotional" and "rational" wards with respect to per capita real estate valuation (the evaluation being twice as high in the rational wards), which resulted from matching them on population density. This may or may not be a serious defect in the study, and Hartmann discounts its importance. He points out that Socialist support does not come primarily from the lowest economic groups and that the socio-economic status of the "rationals" was not as high as suggested by the per capita real estate evaluation, since they were families living in upstairs apartments in the downtown business area. Nonetheless, this difference constitutes an unknown determiner of vote differential. Other difficulties which Hartmann does not mention lie in the tacit assumptions of (a) equal readership of appeals, (b) equal occupational representation, and (c) equal convenience to polls, although the treatment of the data in terms of percentage of increase over the preceding year's vote minimizes any importance of the last two factors named. Another shortcoming of this study, one that is conceptual rather than methodological, was the failure to obtain data of psychological and sociological nature about subsamples of his subjects. Such variables may perhaps be more basic to causing an individual to vote the Socialist ticket than types of propaganda appeal.

The Deren (13) experiment, as yet unpublished, sought to determine the effects of two classes of variables on output in a typewriting situation in which output was expressed in terms of the number of correct strokes per unit time in copying certain materials. The first class of variables may be called "degree of structurization for attaining important ends"; the second had to do with different emotional appeals. The experiment took the following form:

1. A population of 226 students in typing classes of two business schools in Syracuse, New York, first copied a real-estate handbill in the regular class period on days 1 and 2. On the following two days, 3 and 4, the students typed a strongly worded war morale handbill which was to be distributed in various public places in Syracuse.

2. The second phase of the experiment consisted in the delivery, by carefully trained orators, of speeches designed to enhance national morale. Following the numerical scheme used above, days 5 and 6 were occupied with speeches, day 7 was devoted to silent reading of similar material, and days 8 and 9 featured speeches again. Days 10 and 11 were given to further copying of the morale handbill.

The appeal materials were of three basic types and a mixture of two of these types. A survey of propaganda in the press around the time of the experiment showed that most of it could be classified under one or another of the following headings: (a) *Avertive*, calling up emotions of fear, hatred, or rage and making the recipient want to keep the results of losing the war as far from him as possible; (b) *Devotive*, arousing emotions of sympathy, loyalty, or affection toward American or Allied soldiers and making the recipient want to support and share their efforts; (c) *Meliorative*, arousing feelings of altruism toward all mankind and making the recipient want to end the war so that all might enjoy a better world. The first two categories often appeared together in public print.

Appeals of equal length in all three basic categories were specially prepared for the experiment and were judged for strength, or moving-quality, by a panel of judges. They were then revised, further edited, and rejudged. A combination *Avertive-Devotive* appeal was made from the originals of those classes and also judged for strength. In the final forms there were no reliable differences in judged strength among the four sets of material.

The subject population was divided into five groups, one to receive each of the four sets of appeals, and a control.

The comparison of performances on days 1 and 2 vs. days 3 and 4 was to reveal the effects of being structured into a situation (or "national morale system," to use the language of the report) wherein the products of one's work could have important and clearly understandable bearing on maintaining the functioning of social structures important to the worker. The results of this phase of the experiment showed no differences that could be reliably attributed to the change in the structuring of the situation despite the fact that on later questionnaires most subjects reported feeling like working harder when typing the morale handbill than when typing the real-estate handbill. The absence of effect may have been due to the typically high motivation in the typing of classroom exercises (i.e., effects of being in the classroom structure) regardless of the content of what was typed.

The comparisons of scores on the morale handbill typing on days 10 and 11 with days 3 and 4 again showed essentially negative results: no increase in correct typing strokes could be reliably attributed to any of the kinds of appeals or to the appeals taken altogether. Deren concludes with the suggestion that propaganda in the form of emotional appeals has little or no effect in altering behavior output of people in structured stiuations. Inasmuch as this experiment was exceptionally well controlled, this suggestion needs to be taken seriously and investigated with care.

The search for more basic variables received considerable impetus through the study of Lazarsfeld, Berelson, and Gaudet (**25**) of "how the voter makes up his mind in a presidential campaign." In this study, almost sure to become a classic and a point of departure for improved ones yet to come, the stratification techniques of the public opinion pollers were used to sample the population, and some interesting indices were developed in an effort to find important conditions of the population that would be related to voting intention and to voting behavior. Erie County, Ohio, was chosen as the locale for this research because of its convenient size, its freedom from sectional peculiarities and from domination by a large urban center, the possibility of comparing rural with small urban political trends, and because of the fact that its distribution of votes among the various parties had for forty years closely approximated that of the entire country.

Their method of obtaining data was as follows: An initial poll of 3000 eligible voters was made in May of 1940, in which basic sociological data as to occupation, religion, socio-economic status, etc. were obtained. From this original number four groups of 600 each, equated with respect to the mentioned variables, were formed. One was arbitrarily designated a panel, and re-calls were made to its members each month thereafter, the last interviews being taken shortly after the election. The other three groups were controls and were interviewed again only once: one group in July, one in August, and one in October. The panel method thus gave continuous information on changes in voting intention and also provided an opportunity to secure a great amount of additional information about panel members. The control groups served as checks on the effect (which did not materialize significantly) of repeated interviewing on voting intention and behavior.

Important in the present context is one of the several indices and other ways of describing the population that these researchers developed. First, ratings of socio-economic status (SES) were made, using five categories with proportions of the population as indicated: A, 3 per cent; B, 14 per cent; C+, 33 per cent; C−, 30 per cent; D, 20 per cent (A being the highest and D the lowest). Party membership was found to vary

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated
Acc. No

with SES, religious affiliation (Protestant and Catholic only), rural vs. urban residence, and, to only a slight extent, age. From these findings an Index of Political Predisposition (IPP) was developed. This index ranged over six steps from strongly Republican (1) to strongly Democratic (6), and was derived from the values in Table I. To illustrate: a person classified C− in socio-economic status, who is Protestant and lives in the country, would have an IPP of 3, as would a city-dwelling Protestant classified C+ in socio-economic status.

TABLE I

VALUES INCLUDED IN LAZARSFELD, BERELSON, AND GAUDET'S INDEX OF POLITICAL PREDISPOSITION (25)

SES	*Protestant*		*Catholic*	
	Rural	Urban	Rural	Urban
A, B	1	2	3	4
C+	2	3	4	5
C−	3	4	5	6
D	4	5	6	7*

* Sevens were lumped into the sixth IPP step.

Turning to the effects of propaganda, we note that a political campaign may have different kinds of effects. When we speak of "the effect of propaganda" we usually have in mind a change from one position to another on some scale; and whether we get this type of effect depends in part on the scale we use. If our scale is dichotomous, as voting Republican vs. voting Democratic, the only change-in-scale effect possible is outright *conversion*. But another possible effect is what these writers call the *activation effect*, in which the campaign activates the political predisposition of "crystallizers" (those people undecided in their vote intention early in the campaign). Such people were found usually to act in accordance with their IPP's. Still a third effect is called the *reinforcement effect*, meaning the strengthening of original vote decisions, keeping partisans in line, and reducing defections from the ranks. The data on this third type of effect are principally qualitative and were secured in just-before-election interviews. The "pure conversion" effect resulting entirely from Republican propaganda on some campaign issue (for example the third-term issue) must meet these criteria: (1) Voted Democratic in 1936, (2) did not intend to vote Republican before the Democratic Convention, and (3) believed Roosevelt had been a good president. The report states that less than 2 per cent of the total Republican vote in Erie County was attributable to conversions by propaganda on this issue. A few additional people ($N=14$) are reported as being converted by the argument that "Willkie is the champion of the poor," and a very small number—not even reported—of conversions were of voters who conformed to the stereotype of the rational citizen withholding judgment until all arguments were weighed. These people, incidentally, proved to have "weak" IPP's (while they met the three criteria mentioned above, their IPP's were neither Democratic—5 and 6—nor strongly Republican—1 and 2).

This excellent study, making use as it did of the best techniques available at the time for analyzing sociological variables, stands as the prototype of the situational approach to propaganda. Within the limitations of their methods and facilities these researchers even attempted to obtain relevant information on the personalities of the voters, though without much success. This fact suggests their awareness of what we have called the structural approach. One can easily imagine the practical obstacles to an intensive study of personality within a population of any size; the methodological problems which such an undertaking would add to a study like Lazarsfeld's would indeed be formidable. But with sufficient time, a staff of a hundred or more psychologists and sociologists, and a budget increase of several hundred thousand dollars it might be done. These estimates are made merely to suggest the magnitude and difficulties of other than exploratory work of this type.

Having raised the question of the effect of personality characteristics on response to propaganda in an election, we turn to another study of "mass persuasion" which goes further into a consideration of personal motives and their relation to propaganda effects. We are referring to Merton's (**34**) report of Kate Smith's efforts in her third all-day radio marathon to induce people to buy war bonds. Although this study did not give as much attention to sociological variables and was not as well controlled as the Lazarsfeld study, it does show many instances of war bond purchases due to, or at least precipitated by, the urgings of this radio star. It is probably an exaggeration to say that she "succeeded in selling some $39,000,000 of war bonds in the course of one day's broadcasts." It is an exaggeration on conceptual grounds in that there were many factors other than Miss Smith's appeals contributing to the sales, as the report clearly shows, and as Boden (**9**) has also pointed out. It is an exaggeration on methodological grounds because adequate pre-broadcast control data on bond buying tendencies were not secured and because the total sales figure attributed to Miss Smith was an extrapolation from data obtained in New York City where the broadcast was originating. Nonetheless, some very interesting and probably important insights into the workings of a propaganda campaign were uncovered. Let us see how they were obtained.

At eight in the morning on September 21, 1943, Kate Smith began a series of appeals on behalf of the Third War Loan, a series made up of sixty-five separate appearances and extending to 2 A.M. the following day. The procedures for studying Miss Smith's effectiveness contained three major features. First, methods of content analysis (cf. 26) were applied to the scripts used by her in an effort to discover, via the second feature, the focused interviews, the most effective appeals. The focused interviews were conducted by trained personnel in the homes of 100 respondents, lasted three to four hours, and were of the "non-directive type," the respondents being encouraged to report their own foci of attention. The respondents were 75 people drawn from the lists of those who telephoned bond pledges into radio station WABC during the day of the broadcasts plus 25 additional people who had heard the Kate Smith broadcasts but did not respond with a bond purchase. These focused interviews produced data suggesting hypotheses regarding popular images of Miss Smith and other possible motivational factors. These hypotheses could not be tested for validity in the public at large, since data from only 100 intensive interviews were used. Therefore, polling interviews of 978 people in a cross-section of the Greater New York population were conducted to secure additional data. Merton is not clear as to just when the interviews took place. The focused interviews began the day following the broadcast. Half of them were completed within a week and "the bulk of" the others during the week following. No information is given as to the time of the polling interviews. The delay of many of the focused interviews constitutes an additional uncontrolled variable in the observations, for it is well known that original impressions can change greatly with time and depend in unknown ways on personal characteristics (4).

Merton's analysis of interviews and content brings out several characteristics closely relevant to our outline of a structural conceptualization (see pp. 46–47 above). First, there is the organization and content of the persuasive arguments—different appeals evoke different responses and lead to different decisions. Second, the relationship between the propagandist and the propagandee is emphasized, in the form of evidence of the importance of attitudes toward and images of Kate Smith in effecting the decision. Third, the predispositions of the listeners, comprised of a "general orientation" to buy war bonds and a "specific attitude" toward buying more bonds in the Third War Loan drive, are recognized as an important determinant of the effectiveness of the propaganda.

One additional factor which Merton does not recognize as independent of the persuasion situation per se, but which on more careful analysis may be so recognized, is the availability of means for action according to the urgings of the propaganda. The Kate Smith drive allowed the listeners to telephone pledges to their radio station with the understanding that they would later be seen for the actual purchasing transaction. We may be fairly sure that the drive would not have been nearly so successful without this convenient arrangement through which the desired action could be effected (e.g., if people had been asked to run down to the nearest drug store to sign a pledge). We conceive such an arrangement as this not as a part of propaganda itself (i.e., of the independent variable), but as an adjunct to propaganda, a societal intermediary variable, necessary for attaining the objective of the propaganda.

Summary of methodological problems

Now let us point out some of the problems in propaganda research that are illustrated in studies reviewed or referred to above.

1. *Sampling.* Twenty-eight of the thirty-four attitude change experiments reviewed by Murphy, Murphy, and Newcomb (36, chap. XIII) used college or other students as subjects, a limitation on the generality of findings, particularly serious even in exploratory studies when difficulty No. 2 is not met.

2. *Choice of significant independent and intermediary variables.* Practically all studies of attitude change, and Hartmann's action study (20), give most attention to content or its mode of delivery and very little to concurrent conditions within and surrounding the subjects. More fruitful hypotheses for future study would probably have resulted from these experiments, even with the limited population used, had theory and method been more broadly conceived.

3. *Choice of significant dependent variables.* Attitude changes in themselves are of unknown significance with respect to later action. While the difficulties of obtaining action criteria of propaganda effect are appreciated, it seems that in many experiments the question has not even been raised. Flowerman's (18) article suggests a number of measures in addition to attitude that might serve as useful dependent variables.

4. *Adequate measurement and control of independent and intermediary variables.* Most attitude change experiments compare different features of propaganda without effort to control other possibly important aspects of the material. Sims (39), for example, reports different effects on attitudes toward the TVA of pro-utility and pro-TVA propaganda, but his experiment does not control the length, number of themes, quality, or other attributes of the materials, simply selecting some already prepared material at hand. This and other shortcomings also limit Menefee's (33) study of propaganda in the Washington lumber strike of 1935. Such variables making up the main conditions of a propaganda experiment are indeed often hard to measure or control. But their measurement and control are absolutely crucial in answering a question via experimentation—they even determine what question *actually is* being answered by the experiment. Thus, in a research area such as propaganda, where there are many practical difficulties in the way of experimentation and where there is little clear-cut basic knowledge to begin with, the experimenter must be all the more careful to have adequate knowledge of and control over the conditions of his research.

5. *Adequate measurement and control of dependent variables.* The nuances in this problem need fuller appreciation. Attention has already been called to the fact that Merton's (34) interview data were gathered variously over a greater than two-week period following the propaganda, an understandable but regrettable fact. Another noteworthy condition affecting the dependent

variable in some experiments is the attitude of conformity (cf. 1, pp. 298–300, and 43) which operates in group situations to restrict the range of measurements; this condition is likely to vary with different classroom situations (if, indeed, measures *must* be obtained in classrooms). When interviews are the sources of data, special care must be taken to ensure that interviewers are highly and *equally* trained (this applies, of course, to the obtaining of data on intermediary variables as well). In the absence of conclusive evidence of validity in measures used as dependent variables, the conclusions from propaganda experiments should be very carefully and thoroughly qualified, lest misunderstanding and further confusion in propaganda theory and methodology arise. Indeed, a very worthwhile series of experiments could deal directly with the development, refinement, and thorough understanding of dependent variables for propaganda research.

6. *Budget.* One occasionally finds reference in propaganda experiments to the fact that the cost of certain procedures would be prohibitive. No experimenter can be blamed for the high costs of his operations. Caution can, however, be exercised in some cases to keep from spreading funds so thinly that the experimenter loses control over his conditions or is unable to tap variables of possible significance. Consider, for example, the possible differences in results Merton (34) might have obtained had he hired more interviewers to gather data more quickly after the propaganda instead of using fewer interviewers over a longer time, or if Hartmann (20) had spent some of his funds getting better control over sociological variables and less on the leaflets, or if Lazarsfeld (25) had obtained more information on personality variables while spending less money on control-group polling. The upshot of all this is that research in this branch of social psychology is extremely expensive if it is to be done well, and in the absence of truly adequate funds experimenters have some obligation to allocate what they do have so as to yield the most meaningful and valid data.

SUMMARY

We have pointed out that there is as yet no well established body of systematic fact and theory under which the phenomena generally termed propaganda can be subsumed. In an effort to establish this point we have reviewed eighteen discussions of propaganda phenomena. It seemed possible to classify them in the following broad conceptual groups:

1. *Linear conceptualizations*, which seem to view propaganda as a set of forces (or independent variables) inherent in the propaganda materials themselves, which intermingle with other social forces in a sort of ideological tug-of-war causing people to do now one thing and now another (dependent variables). Intermediary or explanatory variables tend to be overlooked.
2. *Psychological conceptualizations*, which emphasize as intermediary variables and as dependent variables the importance of attitudes, values, personality characteristics, or the psychological field in determining the effectiveness of propaganda.
3. *Situational conceptualizations*, which give most attention to the objective facts of social organization as intermediary variables, rather than to values or

attitudes arising from them. Some such treatments were shown to have strong "psychological" admixtures, but the writer's judgment of the relative emphases determined to which category the treatment primarily belonged.

4. *A structural conceptualization* was suggested in which the interdependent functioning of objective social structures and personal characteristics of the propagandees is seen as a complex of intermediary variables determining, and at the same time mediating, the effects of propaganda.

The second major section of the paper reviewed efforts to study propaganda experimentally and pointed out problems in this field of research. Experiments aimed at changing attitudes were treated first and it was concluded (a) that such experiments have on the whole studied the effects of rather superficial independent variables and (b) that change in attitude scale position is of questionable significance as a dependent variable. Four efforts to study the influence of propaganda upon action rather than attitude as the dependent variable were reviewed. These studies highlighted techniques and problems typical of efforts to study propaganda experimentally in "real life" situations. This section closed with a discussion of six important types of problems which must be considered in any kind of propaganda experiment: sampling, choice of significant independent and intermediary variables, choice of significant dependent variables, adequate measurement and control of independent and intermediary variables, adequate measurement and control of dependent variables, and budgeting.

BIBLIOGRAPHY

1. ALLPORT, F. H. *Social psychology.* New York: Houghton Mifflin, 1924.
2. ALLPORT, F. H. An event-system theory of collective action: with illustrations from economic and political phenomena and the production of war. *J. soc. Psychol.*, 1940, **11**, 417–445.
3. ALLPORT, F. H., & SIMPSON, M. M. Broadcasting to an enemy country: what appeals are effective, and why. *J. soc. Psychol.*, 1946, **23**, 217–224.
4. ALLPORT, G. W., & POSTMAN, L. *The psychology of rumor.* New York: Holt, 1946.
5. ANNIS, A. D., & MEIER, N. C. The induction of opinion through suggestion by means of planted content. *J. soc. Psychol.*, 1934, **5**, 65–81.
6. BARTLETT, F. C. *Political propaganda.* Cambridge, Eng.: Cambridge Univ. Press, 1940.
7. BERNAYS, E. L. Manipulating public opinion: the why and how. *Amer. J. Sociol.*, 1928, **33**, 958–971.
8. BIRD, C. *Social psychology.* New York: Appleton-Century, 1940.
9. BODEN, ERIC. Letter to the editor on Merton's "Mass Persuasion." *Int. J. Opin. Attitude Res.*, 1947, **1**, No. 4, 107–108.
10. CHAPMAN, S. H. Review of Merton's "Mass Persuasion." *Int. J. Opin. Attitude Res.* 1947, **1**, No. 1, 121–126.
11. COLLIER, R. M. The effect of propaganda on attitudes following a critical examination of the propaganda itself. *J. soc. Psychol.*, 1944, **20**, 3–17.

12. Cooper, E., & Jahoda, M. The evasion of propaganda: how prejudiced people respond to anti-prejudice propaganda. *J. Psychol.*, 1947, **23**, 15–25.
13. Deren, A. M., & Allport, F. H. The influence of orally presented appeals on civilian morale, defined in terms of war-effort production (as measured by typewriting performance in secretarial schools). Ph.D. dissertation, Syracuse Univ. Library, 1943.
14. Doob, L. W. *Propaganda.* New York: Holt, 1935.
15. Doob, L. W. *Public opinion and propaganda.* New York: Holt, 1948.
16. Droba, D. D. Education and Negro attitudes. *Sociol. soc. Res.*, 1932, **16**, 137–141.
17. Franklin, Muriel, & Allport, F. H. The effect of United Nations versus Axis appeals on morale as shown by the responses on a civilian participation inventory. M.A. thesis, Syracuse Univ. Library, 1943.
18. Flowerman, S. H. Mass propaganda in the war against bigotry. *J. abnorm. soc. Psychol.*, 1947, **42**, 429–439.
19. Edwards, A. L. Propaganda. In P. L. Harriman (Ed.), *Encyclopedia of psychology.* New York: Philosophical Library, 1946.
20. Hartmann, G. W. A field experiment on the comparative effectiveness of "emotional" and "rational" leaflets in determining election results. *J. abnorm. soc. Psychol.*, 1936, **31**, 99–114.
21. Knower, F. H. I. A study of the effect of oral argument on changes of attitude. *J. soc. Psychol.*, 1935, **6**, 315–347.
22. Kornhauser, A. W. Changes in information and attitudes of students in an economics course. *J. educ. Res.*, 1930, **22**, 288–298.
23. Krech, D., & Crutchfield, R. *Theory and problems of social psychology.* New York: McGraw-Hill, 1948.
24. LaPiere, R. T., & Farnsworth, P. R. *Social psychology.* New York: McGraw-Hill, 1949.
25. Lazarsfeld, P. F., Berelson, B., & Gaudet, H. *The people's choice.* New York: Duell, Sloane, & Pearce, 1944.
26. Lazarsfeld, P. F., & Merton, R. K. Studies in radio and film propaganda. *Trans. N. Y. Acad. Sci.*, 1943, Series II, **6**, 58–79.
27. Lee, A. M. The analysis of propaganda: a clinical summary. *Amer. J. Sociol.*, 1945, **51**, 126–135.
28. Lewis, H. B. Studies in the principles of judgments and attitudes: IV. The operation of "prestige suggestion." *J. soc. Psychol.*, 1941, **14**, 229–256.
29. Lindesmith, A. R., & Strauss, A. L. *Social psychology.* New York: Dryden Press, 1949.
30. MacCorquodale, K., & Meehl, P. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, **55**, 95–107.
31. Marple, C. H. The comparative suggestibility of three age levels to the suggestion of group vs. expert opinion. *J. soc. Psychol.*, 1933, **4**, 176–186.
32. McNemar, Q. Opinion-attitude methodology. *Psychol. Bull.*, 1946, **43**, 289–374.
33. Menefee, S. C. Experimental study of propaganda: the Pacific Northwest lumber strikes of 1935. *Soc. Forces*, 1938, **16**, 573–582.
34. Merton, R. K. *Mass persuasion.* New York: Harpers, 1946.
35. Moss, F. A. *Applications of psychology.* Boston: Houghton-Mifflin, 1929.
36. Murphy, G., Murphy, L., & Newcomb, T. *Experimental social psychology.* New York: Harpers, 1937.

37. Parsons, T. Propaganda and social control. *Psychiatry*. 1942, **5**, 551–572.

38. Peterson, R. C., & Thurstone, L. L. *Motion pictures and the social attitudes of children*. New York: Macmillan, 1933.

39. Sims, V. M. Factors influencing attitudes toward the TVA. *J. abnorm. soc. Psychol.*, 1938, **33**, 34–56.

40. Smith, B. L., Lasswell, H. D., & Casey, R. D. *Propaganda, communication, and public opinion*. Princeton: Princeton Univ. Press, 1946.

41. Tolman, E. C. Physiology, psychology, and sociology. *Psychol. Rev.*, 1938, **45**, 228–241.

42. Turner, R. H. Propaganda and the social situation. *Sociol. soc. Res.*, 1943, **27**, 363–372.

43. Vetter, G. B. The measurement of social and political attitudes, and the related personality factors. *J. abnorm. soc. Psychol.*, 1930, **25**, 149–189.

44. Wilke, W. H. An experimental comparison of the speech, the radio, and the printed page as propaganda devices. *Arch. Psychol., N. Y.*, 1934, No. 169.

45. Young, K. *Social psychology*. New York: Crofts, 1944.

46. *The fine art of propaganda*. Institute for Propaganda Analysis. New York: Harcourt, Brace, 1939.

Received March 21, 1950.

A CRITICAL REVIEW OF THE VALIDITY AND RATIONALE OF THE FORCED-CHOICE TECHNIQUE

ROBERT M. W. TRAVERS
Teacher Education Division, Board of Higher Education
New York City

According to a paper by Staff, Personnel Research and Procedures Branch, the Adjutant General's Office (5), the basic idea for the forced-choice technique was developed by Paul Horst with reference specifically to personality scales. Robert Wherry developed a similar idea while working on personality measurement for the Civil Aeronautics Authority, and later the Staff of the Personnel Research Section of the Adjutant General's Office attempted to apply the concept first to the design of personality inventories and later to the problem of rating officers. The latter application resulted in the production of the Army Efficiency Report, which was used for some years for the official rating of officers. More recently the technique has been used by various professional persons and for a number of different purposes. The method of making a forced-choice scale as used by the Personnel Research Section, has been well described by Sisson (3).

The general nature of a forced-choice rating scale is best understood in terms of the procedure which is followed in building it. The following are the essential steps involved:

1. Descriptions are collected of individuals who are at each extreme of the scale to be measured. If the scale is that of effectiveness as a supervisor, descriptions are obtained of the most effective and the least effective supervisors. This procedure partly defines the scale that is to be measured by the final instrument.
2. The descriptions collected are then dissected into a list of small elements. Each one of these elements describes, in essence, a rather specific item of behavior. The complete list should cover all the important aspects of the job, and the number of items covering each aspect should be related in some rational way to the importance of that aspect.
3. Two values are determined for each item listed. One value, the discrimination value, indicates the degree to which the item measures the particular characteristic that is being measured. The other value indicates the extent to which individuals tend to rate others too high or too low on the particular characteristic. This latter is sometimes referred to as the preference index or preference value of the item. Both values are determined experimentally.
4. The characteristics are then arranged in pairs such that the two members of each pair differ in the extent to which they discriminate. Ideally, one of the characteristics in each pair should have a discrimination value of zero and one

should have a high discrimination value. Also, it should be impossible for those who are to use the scale to determine by inspection which one of the two characteristics is the discriminating one.

5. The pairs of characteristics may then be grouped in fours, with each group of four including two desirable and two undesirable characteristics. The main purpose of grouping the characteristics in groups of four seems to be that persons filling out the scale may have a better attitude towards it if the grouping in tetrads is adopted. This seems to be the only advantage of the tetrad over the duad form. A pentad form has also been suggested.

6. Directions are prepared in which the individual is instructed that he is to examine each group of four characteristics and to select the one that is most characteristic and the one that is least characteristic of the person who is being rated. The person filling in the form is required to make these choices.

7. The selection of the items is then validated against an external criterion on a sample that was not used in the original procedure for selecting and pairing the items.

The scoring system is such that the person who is being rated receives a positive score if the item which is most descriptive of him is a discriminating desirable characteristic or if the item least descriptive of him is the undesirable discriminating item. The scores on items may be weighted in some complex manner or the weights may be restricted to the values of zero and unity.

What Does the Rater Do in Completing a Forced-Choice Rating?

The task performed in rating another person on a forced-choice scale is somewhat different from that performed in any other phase of life. Consider, for example, the simplest situation in which two items are presented and the rater is asked to decide which one is most characteristic of the person rated; and suppose that the scale was to be used for evaluating some phase of administrative ability and that the following items constituted the choice offered:

Enjoys making speeches at company dinners
Contributes to company magazine

Suppose that the rater has good evidence to show that the person being rated has a matter-of-fact attitude toward the task of making speeches at company dinners. He does not get any particular enjoyment out of speech-making, but neither does he find the job distasteful. As far as contributing to the company magazine is concerned, he dutifully writes an article every year, as do most executives of the particular company. The situation is confusing for the rater, to say the least, since he is required to compare an item of behavior which varies in degree with one which varies in frequency. Yet the confusion in this instance is much less than that produced by some of the items in the

Army Officer Efficiency Report. In the latter connection, consider the following tetrad of items which appears in the Army scale (WD AGO Form 67-1 Part 2)

A go-getter who *always* does a good job
Cool under *all* circumstances
Doesn't listen to suggestions
Drives instead of leads

In the example, the two words italicized were not in italics in the original version. If these two words are carefully examined in content it will be found that their inclusion makes it impossible to compare these items with the others. Quite obviously a man either is or is not cool under *all* circumstances. It is not an item of behavior which varies in degree. Under such circumstances the rater is asked to perform an impossible task. He cannot say that one of a set of items is more characteristic of the person being rated than any of the other items listed, because some of those listed either are or are not characteristic of the person being rated regardless of the other items in the tetrad.

This confusing situation is mainly a result of the way in which forced-choice scales have been constructed and is not a necessary consequence of the forced-choice technique.

How Can the Choice Be Made Rational?

It is anticipated that one of the necessary changes in the method in order to develop its usefulness as a measuring device will be to modify it in some way so that the rater makes choices only in situations in which choice is rational. One step in this direction would be to include in each tetrad only items of behavior which vary in frequency of occurrence within a given individual or only items which vary in degree. It hardly seems possible to compare behaviors which vary in frequency with behaviors which vary in degree.

The second necessary step in clarifying the method of measurement is in defining what is meant by the "most characteristic" or "least characteristic." Like many terms, these seem clear until they are carefully examined for meaning. For example, when a rater is asked to select from two items of behavior the one that is most characteristic (or most descriptive, or whatever else), he is really being asked to choose the one on which the individual deviates most from the average or from some other standard. If two items of behavior which vary in frequency are under consideration, a comparison of the frequency of one item of behavior must be made with the frequency of occurrence of the other item of behavior in the person rated. This comparison should presumably be made not in terms of absolute numerical frequencies but in terms of some converted score which takes into account the relative scatter of scores of the two variables. The rater might be asked to choose the characteristic on which the individual would have the highest relative standing in his group (which would have to be defined) if all individuals were ranked in order. If the directions could convey this kind of concept it might help to bring meaning into a confused situation.

However, having developed and clarified the psychological concept to this point, the reader will probably have already jumped ahead to ask the question "Why is it necessary to force the rater to choose between the characteristics? Would it not be simpler to ask him to rate the individual on each one of the two characteristics and then examine his ratings to determine which one of the two characteristics was rated higher?" The answers to these questions are in the affirmative, since, if the directions are clear, it is necessary for the rater to assess the individual rated on each one of the items of behavior listed before he can rationally determine "which one is most characteristic" or "which one is most descriptive." The only additional restriction placed by the forced-choice method is that the rating on all the characteristics in a given group must be different, otherwise the forced-choice scoring system cannot be used in its present form.

If the rater is asked to make an assessment of the individual to be rated on each of four characteristics, but with the added restriction that no two ratings within any group of four may be placed at the same point on the scale, the resulting record may be scored in exactly the same way as any of the forced-choice scales at present in use. In other words, the process of forcing a choice is an unnecessary and irrelevant part of the procedure. However, what is left when the procedure is rationalized in the way shown is a scale in which tendencies to overrate or underrate on relevant qualities can be corrected by overrating or underrating on irrelevant qualities. It is assumed in this procedure that, for any given individual, the true rating on the irrelevant items is average and that any deviation in the average ratings on these irrelevant characteristics represents a tendency to overrate or underrate the particular individual who is being rated. The relevant items can be selected empirically, in the first place, to make this assumption acceptable. The procedure is simply that of using ratings on certain characteristics as a suppressor variable to correct for errors in the rating on certain other characteristics. In this form the basic idea of a forced-choice scale seems to have considerable merit and to be one worth developing, though possibly in forms other than the stock-in-trade forced-choice instrument at present in use.

The Validity of the Forced-Choice Rating Scale

The claims for the validity of the technique seem to bear little relationship to the actual evidence.

A primary claim is that "it produces a better distribution of ratings relatively free from the usual pile-up at the top of the scale" (3). The data in Figure 1 were presented to substantiate this claim:

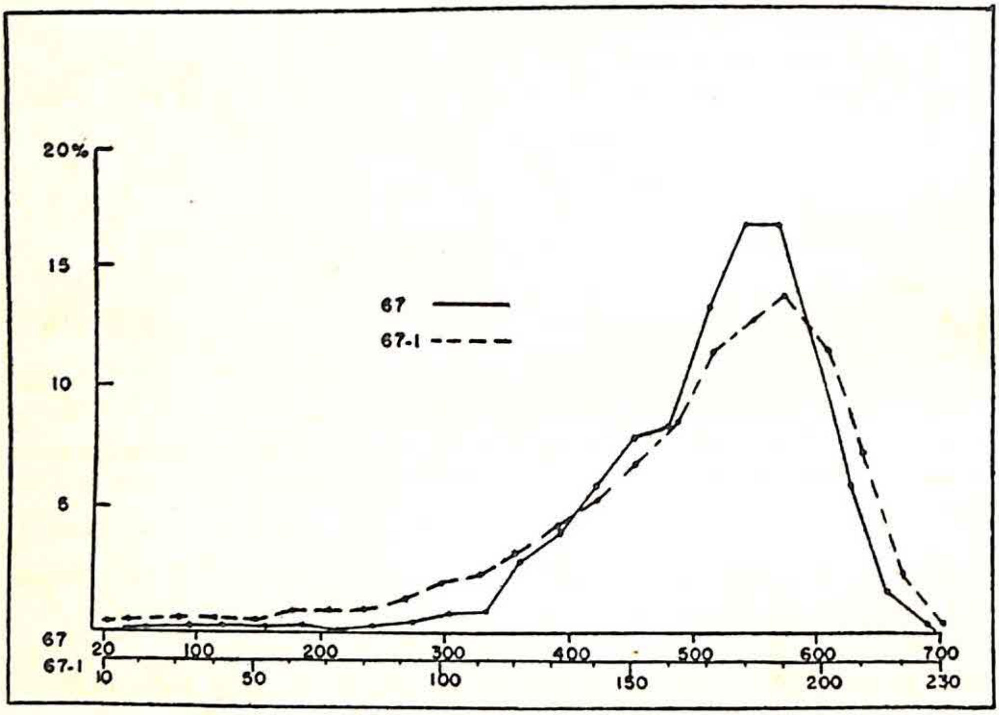

FIG. 1. DATA PRESENTED BY SISSON SHOWING DISTRIBUTION OF RATINGS ON CONVENTIONAL RATING SCALE FORM 67 AND DISTRIBUTION OF RATINGS ON THE FORCED-CHOICE RATING SCALE FORM 67-1 (3, p. 375).

Figure 1 is supposed to compare the distribution of ratings of officers when the old conventional rating scale (Form 67) was used and when the new forced-choice rating scale (Form 67.1) was substituted. It is quite obvious to anyone who examines the above graph that differences in the two distributions do not at all substantiate the claim that the forced-choice rating technique is "relatively free from the usual pile up at the top of the scale." This claim is even more ridiculous when it is substantiated by Figure 2, which presents the contradiction of stating that it is based on data in which range is equated; inspection of the data seems to indicate that the ranges of the two sets of ratings were not equated (data from J. C. Fry).

A second claim is that the forced-choice rating scale as used by the

Army "produces ratings which are more valid indices of real worth" (3). Similar unsubstantiated claims come from Richardson (2), but the apparently spurious nature of the validities presented by him and his group will be discussed later. Claims contradictory to the above are also made by professional personnel who worked on developing the Army Report Form, as is exemplified by the following statement by Staff, Personnel Research and Procedures Branch, the Adjutant General's Office (4):

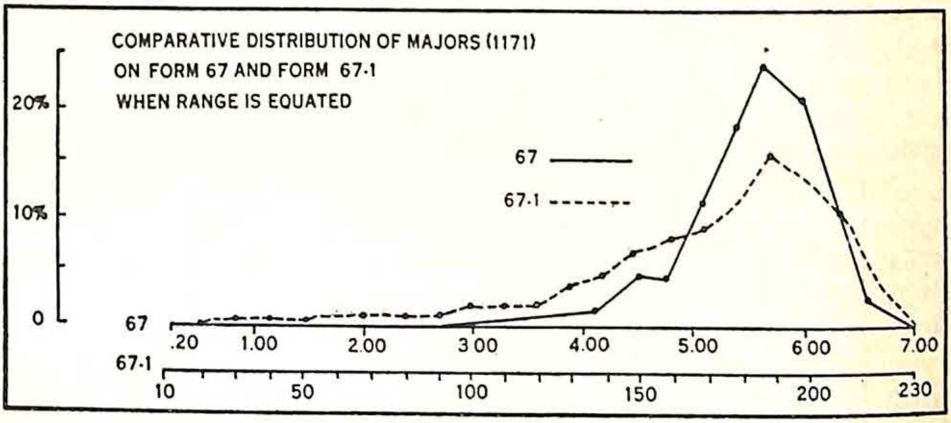

Fig. 2. Data Presented by Fry Showing Distribution of Ratings on Conventional Rating Scale Form 67 and Distribution of Ratings on the Forced-Choice Rating Scale Form 67-1 (1, p. 22).

> A single over-all rating on a 20 point scale, provided it is preceded by a series of specific ratings—using either the paired or unpaired items that have been described—is more valid than either type of specific item alone.

Claims for superiority of the forced-choice rating scale and claims for the superiority of traditional methods coming from the same research offce lead one inevitably to conclude that available data cannot give strong support to the superiority of the one method over the other in so far as Army studies are involved. The following quotation summarizes all the statistical evidence that the present writer has been able to obtain concerning the relative validity of the forced-choice (FCL) and the more traditional type of rating device (RCL) as it is used in the Army Officer Efficiency Report (4):

> The FCL key contains 48 items; the RCL key, 93 items. On the basis of length of rating scale the RCL key would be expected to yield the highest criterion correlations. This is the case for the items scored within the OER-B form, although the difference for the larger group is slight (*r*'s equal .502 for FCL score and .513 for RCL score). When the FCL form and the RCL form are given separately, however, its difference favors the FCL form (*r*'s equal .466 for FCL and .345 for RCL). These findings suggest that when used alone

the FCL technique is superior to the RCL technique. That the two techniques yield similar results when used in combination may well be owing to the forcing of more serious consideration of the individual items when it is known that a choice must be made after rating on each single trait.

Combining the two techniques yields the highest validity for both the item-analyzed group (.540) and the cross-validation group (.562). The gain is considerable when compared with the validity for the RCL technique (.345) or for the FCL technique (.466) used alone, but negligible when compared with the validity for the RCL score obtained when the techniques are used in combination (.513 for the item-analyzed group and .535 for the cross-validation group).

The implication in the above quotation almost seems to be that the presence of forced-choice items has the magical effect of raising the validity of traditional rating procedures. Any such interpretation is obviously nonsense, for it was shown in the same series of studies that a single over-all rating of effectiveness, of the traditional type, was more valid than the forced-choice scale provided it was preceded by a detailed rating system which might be of either the traditional type or the forced-choice type. What the data do show is that the validity of a rating sc is likely to vary to a marked degree with the orientation given to the rater. The data suggest that experimentation with different types of directions may yield much more important results than experimentation with forced-choice scales.

In whatever way these matters may be interpreted, the fact remains that the traditional type of detailed rating procedure and the forced-choice procedure have approximately the same validity when orientation is given; and an over-all rating scale of officer effectiveness may have higher validity than either of these procedures.

Other Validity Studies

Other material presented as evidence of validity is supplied by Richardson (2). The present writer cannot accept the evidence provided by that author and his associates since the procedure involved seems to raise spuriously correlations between assessments of job performance based on a forced-choice scale and an independent criterion of job proficiency. The Richardson technique seems to result in spuriously high validities because of the technique used for refining the criterion. The criterion of supervisory ability is obtained from judgments of those who are in a position to judge the supervisors. Where judges disagree on the rating to be assigned to a supervisor, Richardson discards the case. What is happening, of course, is that those supervisors who are at the extremes of the rating scale tend to be retained while those in the middle of the distribution will tend to be rejected, since those at the extremes can almost inevitably be more accurately judged than those in the middle range. The effect of Richardson's procedure on the distribution of cases on the criterion measure must be considerable since he reports that 51 per cent of the cases were rejected. It need hardly be pointed out that a correlation between the criterion and the forced-choice scale is almost inevitably raised by increasing the relative proportion of cases at the extreme ends of the distribution, particularly when the reliabilities of the criterion rating varied from only "0.4 to a (chance) one of 1.00." It should be noted in this connection that the validity coefficients supplied by Richardson apply to hypothetical populations of cases on which raters agree and do not apply to the population on which the scale was to be used. Such data provide little evidence concerning the validity of the scale when used for the purpose for which it was designed.

Another point to note in connection with the Richardson study is the use of the concept of reliability. While the study shows that a rater is consistent with himself—both throughout one form of the forced-choice scale and from one form to another—in the only case in which different raters used different forms of the rating sheet the reliability of the resulting rating was only 0.69. Also it is not clear whether this reliability was determined on the attenuated or on the original population before cases were discarded. If this particular reliability coefficient, and the others for that matter, were calculated on the basis of the reduced population (with 51 per cent of all cases discarded), they would probably be numerically larger than if the entire population was used.

Other Criticisms of Validity Studies

There is, however, a basic criticism which must be made of all the validity studies which have been reviewed, namely, that they all involve the unsatisfactory process of predicting ratings with ratings. Studies which involve criteria other than judgments must be made and must form the ultimate basis for evaluating the effectiveness of various assessment techniques.

Another claim made by the protagonists of the forced-choice technique is that it prevents the rater from controlling the value of the assigned rating. It is fairly obvious, however, that this is not so. If a rater decides he wants to give a person a higher rating for job proficiency than is deserved, all he has to do is to think of the person who was generally considered to be most proficient in that job and to fill out the forced-choice form for that person. The technique would almost certainly ensure a high score. In a similar way a low score could be assigned at the will of the rater. The method which a rater has to use to control the rating is a little more complex with the forced-choice technique than with the conventional rating technique. It is still an open question as to what fraction of raters will see this loophole in the technique, but loopholes of this kind are likely to become common knowledge.

General Summary

A careful examination of the forced-choice technique as it is used in the Officer Efficiency Report reveals serious defects in the rationale which result in the rater having to make judgments which are strictly illogical. These defects in the technique are not, however, basic and can be remedied.

An examination of the validation studies of forced-choice assessment methods reveals that the evidence does not support some of the claims made for the validity of these procedures. The results show that in studies in which little criticism can be made of the procedure, there seems little to choose between forced-choice and traditional rating. The high validity coefficients secured by Richardson and his associates must be considered to be largely spurious until they are demonstrated to be otherwise.

While validity data published at the time of this review must be considered extremely unsatisfactory and, on the basis of this evidence, the technique does not seem particularly promising. Proper studies need to be made to determine the validity of scales in this area constructed on the basis of an adequate rationale.

BIBLIOGRAPHY

1. FRY, J. C. All superior officers. *Infantry Journal*, 1948, **63**, 21–26.
2. RICHARDSON, M. W. An empirical study of the forced-choice performance report. A paper presented at the 57th annual meeting of the American Psychological Association, Denver, Colorado, 1949. (Mimeographed.) *Amer. Psychologist*, 1949, **4**, 278–279. (Abstract.)
3. SISSON, E. D. Forced choice—the new Army rating. *Personnel Psychol.*, 1948, **1**, 365–381.
4. Staff, Personnel Research and Procedures Branch, Adjutant General's Office. *Comparison of the rating check list (RCL) and forced-choice list (FCL) methods of obtaining ratings.* P.R.S. Report No. 717, 9 July 1946. (Mimeographed.)
5. Staff, Personnel Research and Procedures Branch, Adjutant General's Office. *The forced-choice technique and rating scales.* A paper presented at the 54th annual meeting of the American Psychological Association, Philadelphia, Pennsylvania, 1946. (Mimeographed.) *Amer. Psychologist*, 1946, **1**, 267. (Abstract.)

Received December 27, 1949.

"TASTE BLINDNESS" TO PHENYLTHIOCARBAMIDE AND RELATED COMPOUNDS[1]

WILLIAM C. BOYD

U. S. Naval Medical Research Unit No. 3, Cairo, Egypt

A recent paper by Cohen and Ogden (**7**) raises a number of points which seem to deserve some comment.

1. The authors attribute the discovery that different substances may taste different to different people to Lasselle and Williams (**11, 19**). It is true that these authors reported that creatine is tasteless to some persons and intensely bitter to others. However, there have been no published observations confirming this. When the note by Williams (**19**) appeared in *Science*, I tested about thirty persons without finding one who found the substance bitter. More recently I have tested forty-three persons here, partly Americans and partly Egyptians, and found eight who pronounced the taste "slightly bitter." Nobody seemed to perceive the intense bitter taste which phenylthiocarbamide (PTC) has for those who can taste it. Individuals who taste creatine as intensely bitter must be rare, and of course this taste reaction is not controlled by the same gene which controls the ability to taste PTC.

2. In discussing the possible relationship of the saliva to ability to taste PTC, the authors seem to have overlooked the careful study by Hartmann (**10**), who found that non-tasters could not taste PTC even when it was dissolved in the saliva from strong tasters.

3. Cohen and Ogden imply that sex differences in ability to taste PTC are not significant statistically. In our own work (**4, 5**) differences which were definitely significant were found in several populations. For example, the differences shown in Table I are all statistically significant and in the case of the third population the chances that the observed differences could be due to errors of random sampling are less than one in a million. This sex difference had already been observed by Fisher in England and is shown in his large series of tests, published by Falconer (**8**). It is true that in some populations no significant sex difference in tasting is observed, but we must allow for racial variations, which undoubtedly affect the frequency of the gene for tasting PTC, and also the degree to which the expression of the gene is influenced by sex. Ridell and Wybar (**18**) also refer to the sex difference. Falconer's paper, which also shows the difference, is not mentioned by Cohen and Ogden. In fact, the table presented by Cohen and Ogden and called a "summary of the extensive investigations by interested scientists" does not give the results of Ridell and Wybar, of Fisher, or our own results, and thus

---

[1] The opinions expressed in this article are those of the author and are not to be construed as reflecting the views of the Navy Department.

BIBLIOGRAPHY

1. Astwood, E. B., Greer, M. A., & Ettlinger, M. G. l-5-Vinyl-2-thiooxazolidone, an antithyroid compound from yellow turnip and from *Brassica* seeds. *J. biol. Chem.*, 1949, **181**, 121–130.
2. Boyd, W. C. The assay of blood grouping sera: variations in reactivity of cells of different individuals belonging to groups A and AB. *Ann. N. Y. Acad. Sc.*, 1946, **46**, 927–937.
3. Boyd, W. C. Taste reactions to antithyroid substances. *Science*, 1950, **112**, 153.
4. Boyd, W. C., & Boyd, L. G. New data on blood groups and other inherited factors in Europe and Egypt. *Amer. J. phys. Anthrop.*, 1937, **23**, 49–70.
5. Boyd, W. C., & Boyd, L. G. Sexual and racial variations in ability to taste phenyl-thio-carbamide, with some data on the inheritance. *Ann. Eugen., Camb.*, 1937, **8**, 46–51.
6. Boyd, W. C., & Boyd, L. G. The blood groups and types of the Ramah Navaho. *Amer. J. phys. Anthrop.*, 1949, **1**, 569–574.
7. Cohen, J., & Ogdon, D. P. Taste blindness to phenyl-thio-carbamide and related compounds. *Psychol. Bull.*, 1949, **46**, 490–498.
8. Falconer, D. S. Sensory thresholds for solutions of phenyl-thio-carbamide; results of tests on large sample, made by R. A. Fisher. *Ann. Eugen., Camb.*, 1947, **13**, 211–222.
9. Fisher, R. A. *Statistical methods for research workers.* Edinburgh: Oliver and Boyd, 1932.
10. Hartmann, G. Application of individual taste difference towards phenyl-thio-carbamide in genetic investigations. *Ann. Eugen. Camb.*, 1939, **9**, 123–135.
11. Lasselle, P. A., & Williams, R. J. The identification of creatine. *J. Amer. chem. Soc.*, 1926, **48**, 536–537.
12. Matson, G. A. Blood groups and ageusia in Indians of Montana and Alberta. *Amer. J. phys. Anthrop.*, 1938, **24**, 81–89.
13. Riddell, W. J. B., & Wybar, K. C. Taste of thiouracil and phenylthiocarbamide. *Nature, Lond.*, 1944, **154**, 669.
14. Rikimaru, J. (A study of "taste blindness.") *Jap. J. Psychol.*, 1934, **9**, 87–108.
15. Rikimaru, J. (A study of the individual and racial differences in taste reaction using the same taste stimulus compound.) *Jap. J. Psychol.*, 1934, **9**, 901–921.
16. Rikimaru, J. Taste deficiency of Japanese and other races in Formosa. *Amer. J. Psychol.*, 1936, **48**, 649–653.
17. Snyder, L. H. Studies in human inheritance. IX. The inheritance of taste deficiency in man. *Ohio J. Sci.*, 1932, **32**, 436–440.
18. Snyder, L. H., & Davidson, D. F. Studies in human inheritance. XIII. The inheritance of taste deficiency to di-phenyl-guanidine. *Eugen. News*, 1937, **22**, 1–2.
19. Williams, R. J. "Taste deficiency" for creatine. *Science*, 1931, **74**, 597–598.

Received May 6, 1950.

JOHNSON'S REVIEW OF *WORK AND EFFORT*

M. E. BITTERMAN
The University of Texas

To anyone who has studied Ryan's penetrating analysis of the problems of fatigue and efficiency,[1] Johnson's review of *Work and Effort* must come as a distinct surprise.[2] There is little apparent relation between the book which Johnson describes and the book which Ryan has written.

Ryan's principal thesis is that fatigue and efficiency cannot be satisfactorily evaluated in terms of performance measures alone. Performance, Ryan argues, can be affected not only by the adequacy of working conditions, but also by the effort which the worker devotes to his task. A well-motivated worker may often maintain a high level of output under very inadequate conditions of work as a result of compensatory increases in effort-expenditure, and for this reason conclusions based only on performance measures may be quite misleading. Ryan maintains that considerable attention must be given to the possibility of developing objective indices of effort and devotes several chapters to the status of research in this area. Yet Johnson writes:

> I differ from the author in his presupposing that decrement of performance measures, in general, the effectiveness of an agent of impairment such as prolonged labor, deprivation of sleep, deprivation of oxygen, or dosage of alcohol or drug. For the presupposition denies by implication—but yet effectively—that the living organism is a compensating mechanism, though it really is. . . . Therefore I have to regard part of Dr. Ryan's work as being maldirected (p. 176).

The extent to which Johnson has misrepresented Ryan's position may be judged from the following quotations from *Work and Effort*:

> *On noise.* "A number of different lines of evidence suggest, however, that this adjustment to noise is attained at the expense of additional effort on the part of the subject" (p. 125).
>
> *On temperature.* "One possible interpretation of these results is that the subjects were 'put on their mettle' by the experiment. Perhaps, then, they compensated for the effects of adverse temperature by increased effort, just as they did under the influence of noise" (p. 134).
>
> *On rest periods.* "It is true that most of our knowledge of the effect of rest periods concerns output rather than efficiency. This indicates a gap in our knowledge which must be filled" (p. 151).

---

[1] RYAN, T. A. *Work and effort: The psychology of production.* New York: Ronald Press Co., 1947.

[2] JOHNSON, H. M. Review of T. A. Ryan's *Work and effort: The psychology of production. Psychol. Bull.*, 1950, **47**, 174–177.

On methods of work. "The new pay rates may induce the worker to exert himself more in order to maintain or increase his earnings, and the additional production may not represent a true increase in individual efficiency" (p. 161).

On alcohol. "As Johnson has pointed out, the relationship of accident liability to alcoholic consumption is likely to be very complex. Johnson was referring especially to the tendency toward compensation. It is even possible that a driver could *improve* his driving performance after drinking, because he knows the danger and devotes himself to his driving with extra care and attention" (p. 262).

"Ryan strove for operational definition," writes Johnson (p. 174), and goes on to suggest that this attempt ended in disaster. At the outset of his book, however, Ryan illustrates the limitations of the operational approach and deliberately disavows it (p. 30). Johnson makes much of the fact that capacity cannot be measured directly. Ryan is fully aware of this difficulty but maintains, nevertheless, that the concept is meaningful. He writes:

> Capacity is thus a limit which is never reached in practice, and is a concept closely similar to the concept of *limit* in mathematics. Even though this limit is never reached, and therefore capacity cannot be directly measured, it can be estimated indirectly, and is a construct which is useful in our thinking about human activity, just as the concept of limit is useful in mathematics (p. 31).

Johnson's conception of compensation, by the way, bears not the slightest relation to an operational definition. Adopting his own standards of criticism, we might put to him such holier-than-thou questions as these: Compensation in what person? In what activity? At what time? How measured? The fact is, of course, that compensation, like capacity, is a concept which has meaning even though it cannot as yet be measured directly. And if we refused to give it a place in our thinking until we could measure it, we should never be led to undertake the experiments required to develop appropriate techniques of measurement.

Johnson's struggles with Ryan's formulation of the relation between effort, performance, and capacity afford a strange spectacle. Johnson introduces the assumption that performance may exceed capacity (which is contrary to Ryan's definition) and, after some totally unnecessary symbolic manipulation, emerges with the original assumption, which he only now recognizes to be contrary to Ryan's definition. Forgetting the source of the assumption, he concludes that Ryan's position has been shown to be "very awkward."

Several other brief comments by Johnson express disagreement with points which Ryan did not make. Nowhere, for example, did Ryan imply that the sum of two Poissonian distributions is itself Poissonian.

Nor was Ryan unduly critical of the work of Cobb and Johnson—he merely cautioned the student against unwarranted generalization from Cobb's sample and against *uncritical* application to accident data of techniques which involve the assumption of normality.

Johnson concludes that Ryan's book is "interesting and valuable" (p. 177). It is to be regretted that this evaluation tends to be lost in a wilderness of careless reporting and misdirected criticism.

Received May 1, 1950.

RYAN'S *WORK AND EFFORT:* A REPLY

H. M. JOHNSON
Tulane University

Ryan's definition (d) on his page 96, by which he usually tries to work, exhibits Effort as the ratio between one Power-Output and another. Now by definition, Power $=ET^{-1}$. Given the ratio P/P' between two determinations of Power, and expressing T, T' in terms of the same unit, such as the second, one readily sees that the ratio $P/P' = E/E'$. Hence, Ryan's definition sets forth Effort as a ratio between two determinations of Work-Output, the denominator term P' denoting Work-Output *capacity*. Now if P' was obtained by observation and measurement—that is, operationally—there is nothing in the definition to prevent any determination of P from exceeding it: that is, from turning out more work within a given time than he is capable of turning out. Of course this is merely a conversational necessity; but it is imposed on us by the name that we arbitrarily gave to the denominator-term in the definition.

Of course we could abandon the original definition and substitute a constant value for P', arbitrarily chosen so as to denote an output-rate that no subject could exceed. But the effect would be merely to substitute smaller units of measurement than the one originally employed.

Ryan does sometimes disregard his fundamental definition and substitute one that still implies measurability but which is less precise than the former. In doing so I think he does not improve his presentation.

Perhaps I ought to recall a peculiar form of *compensation:* the subject may shift his mode of attack in any of many ways; he may neglect some demands of his social environment in favor of the task in which his output is being measured. Thus he may maintain a normal output until he is near to collapse. By watching for so-called clinical signs such as these one can learn much about *how* the subject compensates; but in so doing one will contribute nothing to the *measurement* of Effort in Ryan's sense of the term.

I did not say, nor imply, nor even insinuate, that Ryan had undertaken to analyze non-Poissonian distributions into Poissonian components, in the hope of getting generally useful results. My comment was *obiter*, but relevant to some attempts that have been made and published recently.

I do not believe that to ask an author such questions as Dr. Bitterman posed in his twelfth paragraph marks one as being "holier than thou," or holier than anybody else, or even perceptibly holy at all.

Certainly *I* do not refuse concepts places in *my* thinking because they imply non-measurability. I have long held that some existents are *intrinsically* non-measurable; and beginning with 1935 I have been running a series of short articles in the *American Journal of Psychology* on ways of treating such existents without falsifying them in attempts at quantification. I am sorry that they have been disregarded by some who might highly profit by studying them.

Received August 28, 1950.

A NOTE ON LOUTTIT'S REVIEW OF *INDIVIDUAL BEHAVIOR*

ROBERT F. CREEGAN
Ohio University

As a philosopher with a deep interest in the phenomenological movement in general, and in phenomenological psychology in particular, I desire to offer some comments on Professor Louttit's review of *Individual Behavior*,[1] a book which defends the phenomenological viewpoint in psychology. I will also offer a comment on the book itself. I have good reason to agree with the reviewer that few concepts of the particular book are novel. I also agree that attribution of (metaphysical) power to the phenomenal field was unnecessary. I am very much puzzled, however, by Professor Louttit's apparent assertion that the phenomenological concept violates the law of parsimony in psychology.

His argument is that "if the phenomenal field can only be inferred from past behavior, it hardly seems necessary to introduce the entity at all." We may agree that prediction concerns relationships between past behavioral data and future behavioral data. Yet it seems necessary to note a fact which really should be obvious. The data do not themselves predict anything. A psychologist is needed—that is, a thinking observer and systematizer of facts. The data must be reduced to logical order of some sort. Prediction bridges the gap between ordered data of the past and ordered data of the future. Experimentation deals with features of situations which are isolated with reference to ordering concepts. The experimentalist is not concerned with situations in the raw. If this be granted, may we not admit that the "phenomenal field" may be a useful ordering concept?

Phenomenological psychology asserts that, through the discovery of "personal" meanings of behavioral situations, and their "personally" meaningful parts, the data may be ordered in a useful manner for setting up experimental situations of the clinical and the educational, as well as of the more strictly laboratory types. Not every type of experiment is equally useful for every purpose, but the phenomenologist contends that such concepts as "client meanings," "student meanings," and the like, are useful ordering concepts for experiments on motivational development, intellectual development and related trends. Precisely how this would be done is another question. There seems to be no justification, however, for ruling out *a priori* the possibility of ordering the behavioral data in this way, or the possibility that doing so would be fruitful for some research areas in psychology. Ordering concepts do not violate the law of parsimony, unless, indeed, all systematic thinking is unparsimonious. Perhaps science itself is somehow an illegal activity. I do not think, however, that Professor Louttit desires to sponsor the philosophy of irrationalism or primitivism.

---

[1] LOUTTIT, C. M. Review of Donald Snygg & Arthur W. Combs' *Individual Behavior: A New Frame of Reference for Psychology*. *Psychol. Bull.*, 1950, 47, 170–171.

The writer is not disturbed by Snygg and Combs' basic methodology which, as Professor Louttit pointed out, is by no means original, but rather he is disturbed by the fact that their particular phenomenological system has led thus far to few, if any, new empirical conclusions in such a field as motivation. Their contention that "preservation and enhancement of the phenomenal self" is the master motive (or even the sole motive of the normal adult) is strictly "old hat" (I plead guilty to name calling). This theory of motivation seems quite unphenomenological to most of us who are familiar with the literature of philosophical phenomenology. Even Koehler's descriptions of "phenomenal requiredness," which is not always, or perhaps even most frequently, ego requiredness, is more adequate. In conclusion, I would like to raise the question whether ego change is sometimes "required," not in the sense of ego enhancement, but rather in the sense of ego adjustment to, reflection of, or expression of, that which transcends the phenomenal self on dimensions of value experience. Self enhancement may be a function of a field reconstruction which is not a mere response to "phenomenal threats." Or is a self-suspected personal insensitivity one of the chief "phenomenal threats"? The "phenomenal self" of Snygg and Combs seems to be much more threatened by deficiency in personal mastery than by deficiency in personal sensitivity. Perhaps their particular system of phenomenal psychology really is true to "statistically normal" motivation. One reasonable objective for phenomenal psychology, though, is precisely that of placing the statistical norms in a wider perspective! These norms do not reveal all that being a man means, has meant, or may mean. Neither the pathological nor the ideal orders of activity should be omitted by a phenomenological psychology. Some of us think that one of the chief merits of the phenomenological approach is that those who use it will be less tempted, than are "gut psychologists," to imagine that the statistical norms in any way represent viable goals of eduction, or of counseling, in the broadest sense of the terms.

Received April 2, 1950.

LETTER TO THE EDITOR ON PETERS' REPLY TO LEWIS AND BURKE*

C. J. BURKE
Indiana University

In connection with the discussion of chi-square in the July 1950 issue of the *Psychological Bulletin*, the paper submitted by Peters differs in an important respect from the copy to which Dr. Lewis and I wrote our reply. In our discussion on p. 348 we have a reference to Peters' third footnote. In his published criticism of our original paper, the footnote in question was omitted, with the consequence that our remark is left dangling. Further consequences are that our reply to Peters' point 3 on pp. 333–334 of his article is inadequate, since this reply was based upon the missing footnote. The omitted footnote is crucial for our reply and serves better than anything we could write to illustrate Peters' confusions with respect to the chi-square test. The omitted footnote, numbered 3 in Peters' original manuscript, reads as follows:

> There may be needed a re-examination of the applicability of Pearson's formula *vi* to the new conditions for which he did not intend it. He developed the formula on the assumption of freely floating marginal totals, only the total N being fixed. When R. A. Fisher, for the sake of using true values for a narrower universe rather than estimated ones for a broader universe, proposed constant marginal totals with a consequent change in the number of degrees of freedom, he does not appear to have re-examined Pearson's fundamental derivation under these new conditions. With freely floating marginal totals and only a limitation on the total N, a certain system of intercorrelations among the cells arose of which he took account. Under the new conditions a new system of intercorrelations would arise which might very well not be the same. A new derivation with the new assumptions might not come through with the same simple formula *vi*. I am not myself able to carry this through; but invite the attention of expert mathematicians to it.

Our comment on page 348 indicates that the derivation Peters requests is already available in one of the best known statistical sources and that he is mistaken in his notion that the degrees of freedom are changed.

---

* *Editor's Note:* The editor is entirely responsible for the error which caused Drs. Lewis and Burke to criticize a footnote in Dr. Peters' original manuscript which he subsequently deleted from the galley proof of the article. Dr. Peters wrote to the editor before his paper was sent to Drs. Lewis and Burke for comment, and asked that the footnote in question be deleted. He stated that if the paper had gone to press he would take the note out of the galley proof. Through an oversight, the editor did not delete the footnote from the manuscript, nor did he read the galley proof of the article. Galley proofs and subsequent editorial work are now handled in the central office of the American Psychological Association.

Peters' comments on his point 3 are part of the same confusion. Without the footnote just quoted our reply does not cover the points he raises and adequate clarification would require considerable discussion. With the footnote, the discussion as given in our reply is easy and adequate.

Received August 18, 1950.

CHAPANIS, A., GARNER, W. R., AND MORGAN, C. T. *Applied experimental psychology*. New York: John Wiley & Sons, 1949. Pp. v+434. $4.50.

This is the first systematic text dealing with the relations between psychology and engineering—with problems of designing machines so that they are adapted to the characteristics of human beings. It is an important book because it defines the boundaries of this new field and indicates the problems that confront the psychologist who chooses to work in it. The title "Applied Experimental Psychology" was chosen because the research techniques employed most generally in studying equipment-design problems are the traditional ones of experimental psychology. The terms human engineering and engineering psychology are also used in describing the field.

In contrast to most psychology texts, this one is concerned almost exclusively with environmental variables rather than organismic variables. The central problem is, How can man structure his environment so that his job of observing, interpreting, and reacting to it can be done most efficiently? There is no discussion of such familiar topics as selection, training, individual differences, or motivation of workers. Instead, the authors are concerned with the tasks performed by workers and with designing the machines used by them so that the typical individual will perform his job most effectively. Heretofore this point of view has received little emphasis in psychology. Its exposition in the present book reveals many new possibilities for applying psychology.

After the introduction, the reader encounters two chapters on elementary statistics with illustrations chosen from problems of designing machines to suit human characteristics. This brief treatment of statistics will be of little value to the psychology graduate student. Perhaps the chief justification for the inclusion of this section is that design engineers and other readers who are relatively unfamiliar with modern psychology may be impressed with the fact that it is a precise, objective science. The most interesting problem dealt with in this section is how complex man-machine systems can be analyzed so as to identify the major sources of variable error.

The next six chapters deal with visual and auditory perception and with the applications of perceptual data to problems of designing devices for "displaying" information to the human senses. Two more chapters are devoted to the nature of motor skills and to the related problem of designing controls that are adapted to human motor abilities. The sections on "How We See," "How We Hear," and "How We Make Movements" provide a lucid and interesting reorganization of familiar psychological data. The unique sections are those in which the reader is shown how these data can be applied to the design of mechanical and electrical equipment. Much of this material, such as the design of dials and scales, of signalling systems, and of tracking devices, will be new to psychologists. Most of it has become available to psychology as a result of research supported by military organizations during the last decade and carried out at such places as the Air Force's Aero Medical Laboratory, Harvard's Psycho-Acoustic Laboratory, and Johns Hopkins' Institute for Cooperative Research. The last three chapters deal with the arrangement of tools, equipment, and the work place, with work and fatigue, and with the working environment. Much of the material included here is typical of that found in textbooks in industrial psychology and in industrial engineering.

The authors have been very successful in finding ways to display visually the data of psychological experiments. They take seriously their own conclusion that "pictorial visual displays" are easier to interpret than are "quantitative visual displays," and use a total of 196 figures in contrast to only 13 tables as a means of summarizing psychological data. The art work is uniformly excellent and adds greatly to the readability of the book.

Nearly all of the experimental data relative to equipment design problems that were available at the time of publication have been summarized. However, in spite of the inclusion of a number of experiments of questionable importance, the reader is left with the conviction that there are vastly more questions than there are answers in this area of psychological application. A creditable job is done in defining and summarizing these problems, but the picture that the critical reader gains of the level of the systematic and theoretical work in the area leaves him somewhat dissatisfied. This, of course, is more a criticism of the status of work in the area than it is a criticism of the first systematic text covering this new field. In the same vein, it can be pointed out that much of the experimental work on equipment-design problems relates most directly to the design of military equipment, particularly to the design of airplanes, airplane instruments, air traffic control equipment, and radar.

One of the interesting features of the book is its style. The authors originally organized the material for a series of lectures on men and machines at the Naval Post-Graduate School at Annapolis. In amplifying and revising these lectures for publication in book form, they have carried over the notion that they are talking to the reader. They state in the Preface that they have " . . . made deliberate efforts to use small words instead of big ones; short direct sentences instead of long indirect ones; and colloquial expressions instead of academic and technical phrases." A few psychologists of the older school of writing may be somewhat disconcerted to read that "The brain is a mighty fallible instrument" and that "One of the biggest troubles with people . . . is that they get tired." However, this reviewer thinks that most readers will find the style refreshing and will conclude that the material loses nothing of its effectiveness by being presented in small words and short sentences. One result of the style is to make psychological research sound rather easy and quite exciting. This results in part, however, from the fact that the authors have included a minimum of detail in describing experimental techniques and methods. The serious student of experimental psychology will find no precise descriptions of problems, of techniques, or of the exact procedures used in the statistical analysis of data.

"Applied Experimental Psychology" is good general reading. Many psychologists whose primary interest is in another area will want to browse through this book in order to find out what engineering psychology is about. For students with a minimum of training in psychology, such as students majoring in engineering, the book should provide an excellent text. For advanced undergraduates and graduate students in experimental and applied psychology, its chief value will be as a source of supplemental reading. For those who are seeking materials for the introductory course or topics for the experimental course that are related to problems that are of vital importance to people who work in today's technological society, the book will provide a rich source of ideas.

PAUL M. FITTS.

Ohio State University.

CAVAN, RUTH S., BURGESS, E. W., HAVIGHURST, R. J., & GOLDHAMER, HERBERT. *Personal adjustment in old age.* Chicago: Science Research Associates, 1949. Pp. xiii+204. $2.95.

This monograph has three purposes: "(1) to define and to analyze the nature, patterns, and problems of personal adjustment to aging; (2) to present the available facts about old age derived from census data and from a special survey of 2,988 older persons made by the authors; and (3) to describe two instruments for measuring adjustment in later maturity" (p. v). It contains a thoughtful discussion of definitions of aging, of the cultural patterns of old age groups, and of old age adjustment; provides a compact source for census information describing groups past sixty; and presents a number of excellent "sketches" of old people and their modes of adjustment. Although addressed to a heterogeneous audience (research workers, college teachers and their students, social workers, and the general public), in this reviewer's opinion the volume will be read mainly by the first two groups and should be evaluated primarily as a research contribution.

The original data presented in this monograph are of two kinds: (a) information from 499 men and 759 women, describing age trends from the sixties through the nineties with respect to companionship, degree and type of participation, health attitudes, religious attitudes, and adjustment attitudes; and (b) data based on varying N's up to almost 3,000 relative to the construction, validity and reliability, and interrelations of parts of an "Attitude Inventory" and "Adult Activity Inventory," the two being published together as *Your Activities and Attitudes* questionnaire.

The age-trend data provide the best information currently available on the aspects of personality studied, for the years past sixty. Although trends are reported through the nineties, it is somewhat disturbing to see percentages computed in the upper age brackets on N's of 14, 6, 4, and even 2, although attention is called to the sample size in table footnotes. Such reporting is apt to lead to misinterpretation on the part of the non-technical reader, particularly in instances when percentages are given in the text with no N's indicated. At one point (p. 59), the authors themselves mention with reference to "zest for living" that there is "some indication of increased zest for the very last age period," this comment being occasioned by a "reversal" in trend in the very last of eight or nine age categories produced by 33 per cent of three male cases and 100 per cent of two female cases! Although the data presented extend previous knowledge, the value of this report would have been enhanced if age trends for all items had been presented in an appendix. And a more adequate description of the sample (age for age) would have helped the reader judge the extent to which progressive selection had biased the trends reported, an important consideration since most of the findings are based on a 32 per cent return of a mailed questionnaire.

A more general issue relating to age trends deserves special comment. In an early chapter the authors consider at some length the definition of "old age," and recommend definition in terms of some arbitrarily selected percentage of decline. Although such a definition is not without merit, it presupposes an adequate body of information regarding peak years and change during the adult life span, if a reasonable base and percentage are to be fixed upon. Research limited to the age span of this investigation (and many other investigations are similarly limiting their research) not only fails to produce the necessary information regarding the years 20 to 60, but results in a *stage-of-life* psychology of old age, rather than a *developmental* psychology, despite verbal recognition of the latter in this and other publications as the more appropriate view. An example from the present investigation will illustrate the need for a broader age sample. It is noted that practically all the age trends are consistently downward from age sixty on. Yet it makes a great deal of difference, psychologically, whether these trends are continuations of trends originating at thirty or are trends that have *just begun at sixty.*

It is, of course, difficult to evaluate a report written for a heterogeneous audience. Research workers will wish certain information (e.g., in the instance of this report, more information regarding reliability of differences and other statistics) which would tend to clutter up the monograph for the non-technical reader. This reviewer has found the monograph stimulating and useful in courses dealing with the psychology of maturity and old age. It is especially rich in research suggestions, both explicitly stated and implicit in the discussion and findings, in particular in the interrelations among sub-sections of the inventory. The volume merits careful consideration as an important contribution to the growing field of gerontology.

RAYMOND G. KUHLEN.

Syracuse University.

BURROW, TRIGANT. *The neurosis of man.* New York: Harcourt, Brace, 1949. $7.50.

This book represents a recent comprehensive statement of theory and research as products of the Lifwynn Foundation where the author was scientific director. Brief review presents difficulties since the book is an exposition of a new system of behavioral explanation and analysis with its own peculiar terminology. Racial neurosis, phylopathology—an aspect of phylobiology—is thought to be the essential problem rather than the neurosis of the occasional deviant individual. Phyloanalysis, a special form of group and neurodynamic analysis, has discovered two phylogenetically conflicting systems within man that account for his maladaptive behavior. Of more remote biological origin is the tendency for man-in-the-group to react on a group or phyloörganismic basis with a minimum of self-reference and projection and with a maximum of objective group orientation. Of more recent biological origin, associated with the appearance of speech and the ability to utilize symbols, is the development of the "I"-persona, by which is meant "the artificial system of prefabricated affects and prejudices that underlies man's present level of 'normal' feeling and thinking and that leads to inferences that are lacking in biological warrant." Two varieties of attention have been found to be associated with these systems. The one, cotention, occurs when one relates himself directly and in a free functioning way to objects about him. This attitude or attentional trend harmonizes the individual with the phyloörganism. The other variety of attention, ditention, occurs when one is preoccupied with his interests, self-awareness, desires, affects, prejudices and ideas of self-advantage ("I"-persona). Cotention is thought to be phylobiologically integrative, while ditention is disintegrative. The development of the "I"-persona with its ditentive trends is essentially an autopathology that has led to a world or racial neurosis. Since we all share this autopathology, traditional psychiatry and psychoanalysis do little to alleviate the condition. The usual group therapies are similarly irrelevant.

Research records indicate a decrease in breathing rates, a decrease in eye movements, and altered brain wave patterns under conditions of cotention in contrast to records of these responses associated with ditention. Such data are taken to imply neurodynamic bases for these two contrasting phylogenetic systems.

The comprehensive implications of the system for man's welfare are indicated thus: "If insanity, crime, conflict, neurosis and war are to be remedied, . . . ditention and the spurious mood of the 'I'-persona must be made to give way to cotention or to the organism's balanced motivation as a phylic whole." The essential teaching of the book seems to lie in the direction of some kind of self-surrender to the group, the relinquishment of "partitive" self-seeking for individual rights. However, just what final social form or method should be achieved for expressing the ultimate racial non-neurotic norm in healthful, constructive, cotentive cooperation is not clear. The relationship of these implications to a totalitarian state, self-abnegating religions, and the essential philosophy of democracy should have further consideration.

REX M. COLLIER.

University of Illinois.

FLEMING, CHARLOTTE M. *Adolescence.* New York: International Universities Press, 1949. Pp. vii+262. $4.50.

In this compact book on adolescence, the reader will find all of the pertinent facts necessary for an understanding of the maturing individual, without being smothered in a welter of irrelevant and inconsequential detail. This is as it should be, since the text has been written for those who work with and for adolescents.

It is obvious that the author approached the problem with a deep understanding of modern youth. The book is well integrated and reads as if she developed her concept of adolescence and then marshalled the facts in support of her thesis. As a result there is a consistency of viewpoint that is lacking all too often in other texts on the subject. This viewpoint has two aspects: (1) the necessary adjustments to puberty and physiological maturation are determined primarily by cultural rather than biological conditions, and (2) adolescents cannot be treated as a homogeneous group because of the wide range of individual differences and because of the variability of human development. As an educator this leads the author to the logical conclusion that "it is wiser to attempt a progressive adaptation of tuition to individual differences rather than to spend time classifying according to ability and then upon mass teaching of supposedly homogeneous groups." A corollary of this is that education beyond the secondary schools should be widely diversified. In this connection, although Fleming recognizes the usefulness of aptitude tests, she refreshingly does not hesitate to emphasize that the relationship of test scores to achievement is universally low. It is the first time in years that this writer has seen in print the statement that a correlation of .80 is of relatively little value in predicting the performance of an individual on one trait from his score on the correlated aptitude test.

The book is well written in an easy flowing style. Although this leads to a few generalizations to which some might take exception, the book is extremely well documented. References at the end of each chapter run to as high as five pages. After a foreword and an introductory chapter the book is divided into three parts. Part I contains eight chapters and is concerned with the physiological aspect of pubescence and the concomitant changes in needs and attitudes. Part II examines the variety and variability of abilities and capacities of the maturing human being. At this point the unsophisticated reader will likely be more than confused by the statistical concepts employed, but there is a short appendix defining the terms used. The third part is practical and applied. In it the author traces the steps necessary to help the maturing individual in the transition from sociological childhood to adult status and independence.

The question arises as to who will find this book useful. The dedication reads "to teachers, foremen, parents, neighbors, judges, doctors," and "to all who may meet adolescents anywhere." Certainly all of these will find a sound, unemotional, and constructive appraisal of that period between puberty and adulthood. It should also make an excellent text for a one-semester course in adolescence at the undergraduate level; this is especially true if the students are referred to the original sources. Undergraduates will miss the pretty pictures which are gradually replacing the "meat" in American texts, but they will get a well written evaluation of the best research on the subject. For the more mature student of the subject the book's greatest value lies in its extensive biliography of research on adolescence done in the United Kingdom.

RICHARD T. SOLLENBERGER.

Mount Holyoke College.

JOHNS, RALPH LESLIE. *Psychology in everyday living.* New York: Harper, 1950. Pp. xii+564. $3.50.

This is an attractive, readable, brief, simplified text for the first course in psychology. The sentences and paragraphs are short, many concrete cases are cited, and a glossary appears at the head of each chapter. At the end of each chapter, the author provides questions for discussion, lists of relevant films, self-tests, and suggestions for simple experiments, demonstrations, and exercises. These pedagogic devices are utilized effectively to promote understanding and to heighten interest.

Consisting of but 458 pages of text, Johns' is one of a number of brief texts recently published (Johnson, Garrett, Guthrie and Edwards, Brown and Gilhousen) designed for a single semester. Although all the major topics of general psychology are here, yet accuracy and full understanding are sometimes sacrificed for conciseness. Thus we read that strabismus is "imperfect focusing" (p. 181), that in hyperopia the image is focused behind the retina (p. 180), that Binet's co-worker was Thomas Simon (p. 301), that likes and dislikes may be called "brain drives" (p. 87), and that Kuder "correlates intelligence and vocational choice" (p. 312). Tests of mechanical, engineering, business, and art aptitudes are listed as examples of intelligence tests (p. 318), and we are informed that these tests can yield IQ's—not directly, but "by correlation" (p. 318). A questionnaire titled "Have you got what it takes?" and devised by Oliver Hinsdell, who is identified as one who "develops" players for the "screen," is printed as a test of aptitude for motion picture acting (pp. 515–518); no references are made to validity or reliability. We are told that psychology has two divisions, *pure psychology* and *applied psychology*. Instances cited of investigations within the field of pure psychology include the work of Thorndike on transfer of training, Terman on the measurement of general intelligence, and Watson on the behavior of the neonate (pp. 7–8).

The ideal textbook integrates principles and applications: the applications motivate the student and explicate the principles for him; the principles organize his information and provide a rationale for the applications. The "popular" brief text is likely to skimp on basic scientific content relating to the nervous system, learning, intelligence, sensation, and perception in order to make room for "practical" advice on how to be happy though neurotic, how to be popular, how to study, how to choose a vocation, and how to pick a mate. It is cause for gratification that psychology has practical implications; but practice without theory is worthless, if not dangerous.

Can psychologists provide an adequate and meaningful survey of their science in a semester? Useful and informative though this text is, one examines it with this question constantly in mind.

San Diego State College. HARRY RUJA.

STANDEN, ANTHONY. *Science is a sacred cow.* New York: E. P. Dutton & Co., 1950. Pp. 221. $2.75.

"To be read in a hammock on a hot summer afternoon" should be printed on the wrapper of this little book, for it is easy and refreshing reading with its colloquial digs at the pomposities of scientists. We can often laugh with the author and sometimes at him. Much that this court jester at the Shrine of the Sacred Cow has to say is sound, even if not new, and some of it is silly; but all is said with good humor and an amusing style.

Standen, the chemist, knows and tells us on many pages that science means "describing facts," explaining them "by making hypotheses or theories, which must agree with experiments, preferably exactly and quantitatively." What he makes fun of, and rightly so, is "the dreadful cocksureness" of scientists. He admits that there are individual differences. "Though collectively they are a society of mutual backslappers, not all of them go in for scientific evangelism. It is mostly the scientific educators that do." He gleefully quotes a number of purple passages from the exaggerated claims of these educators of what science courses are supposed to do for the students, and then proceeds to tell how deadly dull such courses can be. He is a chemist and ought to know. And the scientists? "As advertising always convinces the sponsor even more than the public, the scientists become sold, and remain sold, on the idea that they have the Key to the Absolute."

Science, Standen quite properly insists, does not yield truth, but only probabilities; it is not "infallible and beyond criticism"; and it has "many important limitations." The best of sciences is physics, and the best things in it "are fully as admirable as the scientists claim them to be." And yet, he says, the teaching of physics is often a messy and unenlightening job. The remaining sciences "can all be arranged in order, going from fairly good through mediocre to downright bad": biology, psychology, and the social sciences.

Biology, Standen tells us, is a great collection of facts with a "vast mass of analogies" and sweeping generalizations. He cleverly puts his finger on the unfilled gaps between species.

Psychology, the more important part in the proper study of mankind, "comes from the Greek word meaning soul, and would be clearer, but not more elegant, if we called it soulology. There is nothing mysterious in saying that man has a soul. Think of yourself, are you your body? It is "your" body: the "you" that has the body is your soul. . . . The body has weight, the soul does not: this is a perfectly clear distinction between the two." Simple, isn't it? But the poor benighted psychologists do not see it this way because they "can't ask the right questions, and they have given up (to a greater or less extent) the best way we have of understanding men." The real trouble with psychologists is that "being fundamentally muddle-minded" and having noticed the "close connection between the body and the soul they suppose that they are the same thing. . . . All of them pride themselves on their ability to reason, and most of them have a tendency to be a little humorless. . . . The less scientific they are the better they are doing." Which best-seller may he be talking about?

According to Standen there are plenty of people to go to beside psychologists "if you really want to understand human beings." They are the novelists and poets. He does not realize that he is quoting from some textbooks. If he should ever find out, he will have to abandon his own cocksureness and ask himself why some psychologists have said precisely the same thing. He might then realize that it is dangerous even for a chemist, "first cousin" to the physicist, to pass judgment on workers outside his own field.

The author is far from belittling all that has been done in psychology "which has a number of fine achievements to its credit: treatment of war-shocked soldiers, and far better understanding of some kinds of mental illnesses . . . aptitude testing, job analysis, psychology of advertising (sic), attention-getting techniques, and methods of 'motivating' other people. But for all the usefulness of these, the really important questions in human life are hardly touched upon at all by psychologists." You will get little help from psychologists if you want to know about "really important human qualities, such as goodness, or holiness," or the various virtues. I have to admit that the author's aim is getting better.

The social scientists come off worst. Many of them "have an aggressiveness and cocksureness hard to match, even among the psychologists." They are as busy as the bees collecting facts so that they can do some correlating. In order to be scientific they frame hypotheses and proceed to test them. "This can always be done and it makes anything at all sound wonderfully scientific. If you want to find out whether your pencil is in the top drawer of your desk, you first frame the hypothesis. This is always verbally possible, but it reduces 'hypothesis' to such a miserably feeble thing that all the high-tooted virtues of the scientific method have been squeezed out of it." A good shot.

There is one curious omission in this catalog of the sins of scientists from the exalted physicists through the mediocre psychologists to the lowly sociologists: Nowhere is the distinction drawn between scientists and technicians. In fields with well established techniques and beautifully equipped laboratories it seems easy to become a respected technician turning out article after article and perhaps even a textbook, without knowing too much about the nature of science. The need for an understanding of what science is about is greater in the less developed fields. It is barely possible that there is more striving for such an understanding in the "mediocre" and "bad" sciences than in the "admirable" ones. At any rate, mere technicians posing as scientists are to be found at all levels.

Writing a book—or a review of it—is as revealing as a projective test. Although there is no mention of Thomas Aquinas, his spirit pervades the book like a faint and delicate aroma. Witness the concern with truth, virtues, the soul, and Plato's higher knowledge. And rising to a climax towards the end of the book, the author declares: "The first purpose of science is to learn about God, and admire Him, through His handiwork." The reviewer replies: The purpose of science is to describe nature without God. In doing so science has definite limits which are fixed by its methods. If scientists want to concern themselves with God they can do so only outside those limits.

KARL F. MUENZINGER.

University of Colorado.

GUILFORD, J. P. *Fundamental statistics in psychology and education.* (2nd Ed.) New York: McGraw-Hill, 1950. Pp. xiii+633. $5.00.

Unlike many revisions of popular textbooks, Guilford's second edition stands out as a completely rewritten text. Although some of the chapter headings look familiar, the extensiveness of topical coverage is considerably greater than in the first edition.

The first eight chapters of the book are devoted to a thorough yet clear presentation of descriptive statistics designed primarily for the introductory course. Following this, three chapters devoted to statistical inference represent considerable improvement over the superficial coverage of this highly important topic in the first edition. Nevertheless, this portion of the book still leaves much to be desired in view of the increasing use in psychology and eductaion of such techniques as analysis of variance.

The remainder of the book supplies a much needed, up-to-date discussion of the logic and statistical theory underlying measurement and prediction. It is here that Guilford is at his best. Articles concerned with test reliability and validity, scaling procedures, and prediction problems have mushroomed forth during and since the war to an extent which makes it exceedingly difficult for most psychologists to keep up with the progress being made. The consistent and clear treatment of these topics should prove most useful to research workers and advanced students who wish an orientation in modern test theory.

Several minor points may prove disturbing to some readers. Snedecor's F Table is *not* applicable to the simple case of testing the differences between two standard deviations unless the levels of significance are adjusted from .05 and .01 to .10 and .02, respectively (p. 233). To test the homogeneity of the "within-set" variances in analysis of variance the highly questionable procedure of making an F test for the difference between any two sets of estimates is advocated (p. 257). No mention is made of Bartlett's test of homogeneity or similar tests useful for this purpose. The discussion of analytic procedures in a two-way analysis of variance problems is superficial and somewhat misleading, particularly in regard to the appropriate variance estimate to be used in the denominator of the F test when the interaction mean square proves significant.

Apparently in order to simplify the discussion of statistical inference, Guilford has used the term "*t*-ratio" indiscriminately for both the critical ratio test and Fisher's *t* test. This may lead to some misinterpretation by the reader, as for example on page 278: "When there is 1 degree of freedom in a contingency table, chi square is equal to t^2." This is true if *t* is thought of as a normally distributed variable. On the other hand, if *t* has Fisher's "*t* distribution," the relationship is an asymptotic one as the sample size approaches infinity.

The generous use of timely data from recent psychological research to illustrate and clarify the application of statistics throughout the book is quite refreshing. In addition to a glossary of the more important symbols commonly used in statistics and a section consisting of simple proofs and derivations, the appendices contain four tables not in the first edition. The thoroughness of revision and clarity of presentation should make this book highly useful as a basic text for both introductory and advanced courses in statistical methods.

WAYNE H. HOLTZMAN.

University of Texas.

BOOKS AND MATERIALS RECEIVED

ABT, LAWRENCE E., AND BELLAK, LEOPOLD. *Projective psychology. Clinical approaches to the total personality.* New York: Knopf, 1950. Pp. xvii+485+xiv. $6.00.

BALFOUR, M. C., EVANS, R. F., NOTESTEIN, F. W., AND TAEUBER, IRENE B. *Public health and demography in the Far East.* New York: Rockefeller Foundation, 1950. Pp. vi+132.

BRAYFIELD, ARTHUR H. (Ed.) *Readings in modern methods of counseling.* New York: Appleton-Century-Crofts, 1950. Pp. xii+526. $5.00.

BROWN, WARNER, AND GILHOUSEN, HOWARD C. *College psychology.* New York: Prentice-Hall, 1950. Pp. xviii+507. $3.75.

COLE, LAWRENCE E., AND BRUCE, WILLIAM F. *Educational psychology.* Yonkers-on-Hudson: World Book Co., 1950. Pp. xvi+768. $4.50.

CORNELL, F. G., MCLURE, W. P., MILLER, V., AND WOCHNER, R. E. *Financing education in efficient school districts.* Bureau of Research and Service, Coll. of Education, Univ. of Illinois, 1949. Pp. 165.

FARNSWORTH, PAUL R. *Musical taste: its measurement and cultural nature.* Stanford, Calif.: Stanford Univ. Press, 1950. Pp. 94. $1.50.

DORCUS, ROY M., AND SHAFFER, G. WILSON. *Textbook of abnormal psychology.* (4th Ed.) Baltimore: Williams and Wilkins, 1950. Pp. xvi+717. $5.00.

Ferm, Vergilius. *A history of philosophical systems.* New York: Philosophical Library, 1950. Pp. xiv+642. $6.00.

Fordor, Nandor, and Gaynor, Frank. *Freud: dictionary of psychoanalysis.* New York: Philosophical Library, 1950. Pp. xii+208. $3.75.

Fromm-Reichmann, Frieda. *Principles of intensive psychotherapy.* Chicago: Univ. of Chicago Press, 1950. Pp. xv+246. $3.75.

Guilford, J. P., and Lovell, Constance. *Advanced statistical exercises.* Beverly Hills, Calif.: Sheridan Supply Co., 1950. Pp. ii+110. $2.00.

Harrower, Molly R. (Ed.) *Recent advances in diagnostic psychological testing.* (American Lecture Series, Publication No. 81.) Springfield, Ill.: Charles C Thomas. Pp. x+120. $2.75.

Hartley, E. L., Birch, H. G., and Hartley, Ruth E. *Outside readings in psychology.* New York: Thomas Y. Crowell, 1950. Pp. xiii+875. $2.75.

Hellersberg, Elisabeth F. *The individual's relation to reality in our culture.* Springfield, Ill.: Charles C Thomas, 1950. Pp. x+128. $3.25.

Gilbert, G. M. *The psychology of dictatorship.* New York: Ronald Press Co., 1950. Pp. viii+327. $4.00.

Homans, George C. *The human group.* New York: Harcourt, Brace, 1950. Pp. xxvi+484. $6.00.

Keller, F. S., and Schoenfeld, W. N. *Principles of psychology.* New York: Appleton-Century-Crofts, 1950. Pp. xv+431. $4.00.

Keys, A., Brozek, J., Henschel, A., Mickelsen, O., and Taylor, H. L. *The biology of human starvation.* Minneapolis, Minn.: Univ. of Minnesota Press, 1950. Vol. I: Pp. xxxii+763; Vol. II: Pp. viii +620. $24.00.

Krout, Johanna. *Symbol Elaboration Test (S.E.T.): the reliability and validity of a new projective technique. Psychological Monographs* No. 310 (Vol. 64, No. 4). Washington, D. C.: American Psychological Assn., 1950. Pp. vi+67. $2.00.

Leeds, Carroll H. *A scale for measuring teacher-pupil attitudes and teacher-pupil rapport. Psychological Monographs* No. 312 (Vol. 64, No. 6). Washington, D. C.: American Psychological Assn., 1950. Pp. iv +24. $1.00.

Meehl, Paul E. *Minnesota Multiphasic Personality Inventory research results for counselors.* St. Paul Dept. of Education, August, 1950. Pp. 42.

Morgan, Antonia Bell. *Modern English Usage Test.* Washington, D. C.: Aptitude Test Associates, 1950.

MURDOCK, G. P., FORD, C. S., HUDSON, A. E., KENNEDY, R., SIMMONS, L. W., AND WHITING, J. W. M. *Outline of cultural materials.* Behavior Science Outlines, Vol. I. New Haven, Conn.: Human Relations Area Files, Inc., 1950. Pp. xxiii+162. $2.00.

PROTHRO, E. T., AND TESKA, P. T. *Workbook for psychology: a biosocial study of behavior.* Boston: Ginn and Company, 1950. Pp. vi+113. $1.40.

TOBINSON, MARGIE. *More than medical. Veneral disease control—a psychological and social problem.* Portland, Ore.: E. C. Brown Trust, 1950. Pp. 36. $.25.

SCHOFIELD, WILLIAM. *Changes in responses to the Minnesota Multiphasic Inventory following certain therapies. Psychological Monographs* No. 311 (Vol. 64, No. 5). Washington, D. C.: American Psychological Assn., 1950. Pp. vi+33. $1.00.

STEWARD, JULIAN J. *Area research: theory and practice.* New York: Social Science Research Council, 1950. Pp. xix+164. $1.50.

WOLFF, WERNER (Ed.). *Values in personality research.* Personality—Symposium No. 1, April, 1950. Pp. 74. $2.00. *Projective and expressive methods of personality investigation.* Personality—Symposium No. 2, April, 1950. Pp. 76. $2.00. New York: Grune & Stratton.

WOLFF, WERNER. *Values and personality.* New York: Grune & Stratton, 1950. Pp. x+239. $4.75.

Control of radiation hazards in the atomic energy program. United States Atomic Energy Commission. Washington, D. C.: U. S. Government Printing Office, 1950. Pp. xi+230. $.55.

In-service preparation for guidance duties. Office of Education, Federal Security Agency. Washington, D. C.: U. S. Government Printing Office, 1950. Pp. iv+48. $.30.

Memorandum on university research programs in the field of labor. Committee on Labor Market Research. New York: Social Science Research Council, 1950. Pp. vi+64.

Procedures for evaluating research personnel with a performance record of critical incidents. Manpower Branch, Human Resources Division, Office of Naval Research. Pittsburgh: American Institute for Research, June 1950. Pp. iv+42.

Vol. 48, No. 2 March, 1951

Psychological Bulletin

A CRITICAL REVIEW OF LATENT LEARNING AND RELATED EXPERIMENTS[1]

DONALD THISTLETHWAITE

Western Reserve University

INTRODUCTION

Latent learning experiments have been the object of considerable interest and controversy in recent years. Careful inspection of this literature strengthens the belief that a sufficient number of experimental studies have now been accumulated to enable us to state with some specificity what conditions favor and what conditions do not favor such learning. Such a statement would aid us in interpreting apparently contradictory results, and it would facilitate the design and planning of research aiming to clarify the relationships between latent learning experiments and the various theoretical issues with which they have been associated.

The two principal theoretical issues involved are: (1) Is reinforcement a necessary condition for learning? and (2) Should a distinction be made between learning and performance? Current interest in latent learning is largely attributable to the opposite and mutually exclusive positions of reinforcement and nonreinforcement learning theorists on the first question. The former group of theorists (Thorndike, Hull, Spence, *et al.*) have postulated that certain rewarding "after-effects" or reinforcements are a necessary condition for learning. Tolman, Leeper, White, and others have argued cogently, on the other hand, that the latent learning data constitute strong evidence that reinforcement is not required for some learning. This issue is by all odds the most important one raised by this series of experiments, and it has been clearly restated by recent investigators (**30, 47**).

---

[1] The present paper is a portion of a dissertation presented to the Graduate Division of the University of California at Berkeley, in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to express his sincere appreciation to Professors Edward C. Tolman, Warner Brown, Robert W. Leeper, and Richard S. Crutchfield for generous advice and criticism.

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated ....

Whereas there is agreement on the fact that one group of learning theorists regard reinforcement as necessary for learning, and another group of learning theorists do not, no similar unanimity of opinion exists as to the bearing of latent learning upon this issue. To mention but a few of the participants in this discussion, Tolman (65) and Leeper (38), as well as Seward (53), have interpreted the latent learning data as crucial in evaluating these two theoretical positions. On the other hand, Meehl (46) argues that "even the Strong Law of Effect is . . . compatible with the latent learning experiments" (p. 70). Similarly, Postman writes:

> Crucial experimental tests to arbitrate between these two interpretations of more complex forms of learning [i.e., between reinforcement and non-reinforcement interpretations] still seem to be lacking. For the time being, therefore, acceptance or rejection of the hedonistic postulate must remain a matter of philosophical preference (50, p. 501).[2]

We shall attempt to show that the results of recent experiments are by no means so inconclusive on this issue and that within the series of experiments on latent learning there are rather crucial experimental tests which enable us to reject, or at least to regard as highly implausible, one of these alternative types of interpretation.

The second issue distinguished above is less controversial, and although a number of theorists are agreed that the answer should be in the affirmative, recent statements by others suggest that the distinction between learning and performance has either not been accepted (44) or has not been found necessary (46). This distinction was first suggested by Lashley (36) and Elliott (12) and later stressed by Tolman (64) and Leeper (38). Hull's acceptance of it is embodied in his distinction between habit strength and effective reaction potential. Both Hull (27) and Spence (56) have stressed that habit strength may not be exhibited for some period in discrimination training. The issue between *these* theorists is not whether learning may occur which is not manifest in performance. Both groups admit that it may. The issue is rather one as to the cause or source of learning, i.e., whether there are any causes of learning other than reinforcement. On the other hand, McGeoch (44) has questioned the validity of the distinction, arguing in effect that such a distinction is undemonstrable. And more recently, Meehl (46) has formulated the law of effect and defined "trans-situational reinforcers" so that reinforcers strengthen *responses.* Such a formulation would make difficult any distinction between learning and performance. We shall review many instances in which learning is unexpressed, or only partially expressed, in performance and which appear to justify empirically the distinction in question.

---

[2] A similar scepticism concerning the testability of the law of effect has been expressed by Meehl: "So long as the Strong Law of Effect is stated qualitatively and does not explicitly mention amounts and times, it would be admittedly difficult to design an experiment in which it could be refuted" (46, p. 70). Yet, elsewhere in the same paper this writer argues that the same law *is* testable: "Such a statement (i.e., the Strong Law of Effect) *can clearly be false to fact*, either because no such trans-situational reinforcer can be shown to have been present, or because the same learning can be shown to be possible without it in the present setting" (p. 65). (Author's italics.)

CLASSIFICATION OF LATENT LEARNING EXPERIMENTS

Four principal types of latent learning experiments may be distinguished:

1. Type I, or the Blodgett variety of latent learning experiment, consists in giving the rats a series of unrewarded, or slightly rewarded, trials in a maze. A relevant goal object is then introduced and additional trials are given (2, 11, 19, 48, 51, 54, 70, 73, 75).
2. Type II is characterized by the procedure of permitting the animals to explore and live in the empty maze for brief periods of time. Subsequently a relevant goal object is introduced into the maze, and a single test trial or a series of test trials are given (5, 9, 15, 28, 35, 38, 53).
3. Type III consists in giving rats which are satiated for food and water a series of trials in a maze the pathways of which contain the goal objects for which the animals are satiated. After this series of trials the rats are made hungry or thirsty and given additional free trials in the maze (30, 43, 47, 59, 62, 63).
4. Type IV involves giving thirsty or hungry rats a series of trials in a maze containing relevant and/or irrelevant goal objects. The animals are then satiated for the formerly desired goal object and deprived of the previously undesired goal object. One or more trials are then given in the practiced maze. Many investigators have failed to find evidence of latent learning with this type of design (13, 14, 31, 33, 41, 60, 71), whereas positive results have been reported by others (8, 10, 17, 38, 61, 72).

Obviously these brief descriptions must be elaborated considerably before they can be said to state the conditions under which latent learning occurs or fails to occur. We shall shortly consider each of the above classes of latent learning experiments in greater detail. But first we must state the criteria according to which some learning in these experimental situations is said to be "latent."

Historically, the term "latent learning" has been used in two senses: (1) to refer to a learning not manifest in performance scores, and (2) to refer to a learning which occurs when the incentive is irrelevant to the existing drive (or drives). Blodgett, who coined the term and used it in the first sense, writes:

> Do these sudden drops in errors which come after the introduction of reward indicate that something to be called *latent* learning developed during the non-

> reward period—a latent learning which made itself manifest after the reward had been presented? Obviously, if this is to be answered in the affirmative, one must show that the drops which occurred immediately subsequent to the reward were larger than "normal," that is, that they were bigger than drops in the control curves (2, p. 122).

Blodgett suggested two alternative criteria for determining whether such error decreases were larger than normal:

1. Greater error decreases in a non-reward group of rats upon being rewarded than any day-to-day decreases in a consistently rewarded group.
2. Greater error decreases in a non-reward group of rats upon being rewarded than in the consistently rewarded group's error decreases interpolated so that the number of errors at the beginning of the critical reductions were equal for both groups. This interpolative procedure has been described as one in which "reduction in error score for the non-reward group was compared with reduction in error score *at an equivalent ordinate* for the control" (3, p. 185).

Some investigators have questioned whether the above criteria are sufficient. Thus it has been argued that decreases in the degree of cul entries (51), or possible changes in the respiration, pulse rate, etc., of the rat (28) might provide manifest indices of the acquisitions taking place. In experiments of the Blodgett type, the question of whether there are *some* manifestations of *some* learnings during the latent period is comparatively unimportant. It is not necessary that there be no evidence of a tendency toward error reduction during the unrewarded (or slightly rewarded) trials in order for the unmanifested learning to qualify as "latent learning." The question of whether the abrupt increase in efficiency of running the maze upon the introduction of food is significantly greater than the comparable improvements of a control group is a statistical question. If it is in any given experiment, then that experiment may be interpreted as yielding positive evidence of latent learning.

Previously Spence and Lippitt (59) and Walker, Knotter, and DeValois (72) described their results in terms of the relationships of relevance or irrelevance between drives and incentives (the second of the two meanings listed above). However, the criteria according to which this relationship of "relevance" is to be determined have not yet, it seems, been explicitly stated. The question of the relevance of drives and incentives to each other is a complex one. However, knowledge of the procedures of deprivation (or experimental satiation, whichever is the case) to which the animal has been subjected and descriptions of his behavior are sufficient to determine the relationship between drives and incentives. If a rat is deprived of an object normally desired, or required for self-maintenance, and if the rat then selects pathways leading to this goal object and manifests consummatory responses toward it, then the goal object, or incentive, is clearly *relevant* to the animal's drive (or drives). If a rat given access to an incentive until it no longer manifests consummatory responses toward the incentive, then given a series of free trials on a multiple-path maze with which it is familiar, does not select any more frequently than chance the path leading to the incentive, and, if having selected the path leading to the incentive, does not manifest consummatory responses toward it, then the incentive is clearly *irrelevant* to the animal's drive (or drives).[3] On this basis, we shall speak of "irrelevant-incentive learning," i.e., learning of the location of, and means of arriving at, an object or incentive which is irrelevant to the drive (or drives) existing at the time of the learning.

In what follows, the term "latent learning" will be used to refer to the learning involved in experiments of types I and II as classified above, and the term "irrelevant-incentive learning" will describe the learning involved in types III and IV.

EXPERIMENTAL STUDIES OF LATENT LEARNING AND IRRELEVANT-INCENTIVE LEARNING

I. *Experiments Utilizing Unrewarded Trials*

The relationships between drives and incentives may be altered either by varying the motivational state of the animal or by manipulating the incentives present in the maze situation. In the kind of experiment designated as type I it is the reward value, or incentive, which is manipulated. Simmons (54) reported the first experimental evidence suggestive of latent learning in this type of situation in 1924. Shortly thereafter, Blodgett (2) undertook a more systematic study of the phenomenon and introduced the term "latent learning." Blodgett used as a food reward a mash saturated with water; the animals were given no other water. This means that the animals were actually run under both hunger and thirst motivation. One group (the control) was fed in the food box for three minutes after each trial. A second group was treated like the first in every respect except that it was given no reward at the end of the maze until the third day, while a third group was given no reward until the seventh day. Blodgett found that the animals when suddenly rewarded improved greatly in performance, their errors dropping sharply to a level comparable to that of a control group which had been rewarded from the beginning. When evaluated by the second of the two criteria listed above, these error decreases were very significant for the three-day group and bordering upon statistical significance for the seven-day group.

---

[3] Postman, in criticizing the "irrelevant reward" experiments of Thorndike and his associates, writes: "Relevance and irrelevance are terms which refer to the *experimenter's* interpretation of the situation. . . . Relevance is a normative term reflecting the experimenter's judgment" (50, p. 515). Whereas this criticism applies to the experiments of Lorge (42) in which an *instrumental* act brought reward, it does not represent a damaging criticism of the above operational and behavioral definition of relevance. Most observers would probably grant the intersubjectivity of terms defined in this way. The concept of "irrelevant-incentive learning" should not be confused with learning by means of "irrelevant reward" (Lorge), since the latter refers to a learning which occurs *after* an originally incidental act is made relevant to the animal's drives.

Herb (19) found that when the food reward was introduced into the blinds instead of at the end of the true path, her rats exhibited a significant increase in the number of cul entries after the introduction of reward. This suggests that the rats were acquiring during the unrewarded period some kind of readiness which was functional in leading to the blind alleys as well as to the goal box at the end of the true path.

Latent learning of this kind appears to occur most readily when the following favorable conditions are present:

1. *There is a complex maze or task to be mastered.* Blodgett used a six-unit T-maze. Tolman and Honzik (70) used a 14-unit T-maze and obtained, following the introduction of reward, sudden drops in error and time scores which were steeper and of greater statistical significance than Blodgett's. This difference in results may be merely a psychometrical difference arising from the use of apparatuses yielding different ranges of scores. When the test for latent learning is made by comparison of the decreases in error and time scores of experimental and control groups, as in this type of design, any increase in the sensitivity of the measurements involved should bring out more clearly any genuine differences which exist as a result of the experimental procedures.

2. *The reward introduced is highly desired by the animals or has been associated with an incentive which is highly desired.* Williams (75) demonstrated that a goal box which had been closely associated with food could, under certain conditions, function as an incentive. Halfway through the training she introduced a black and white goal box (in which her rats were accustomed to find food) at the end of the maze. Although no food was present in the box following the substitution of this box for the strange, empty one to which the rats had run during the "unrewarded" trials, there occurred a decided improvement in maze performance. Williams originally suggested that "the extent to which the learning curve drops when a reward is introduced may be taken as a measure of the value of the reward" (75, p. 33). Similarly Elliott (11) showed that the effect produced by the substitution of a new goal object was dependent upon the reward value of this object as determined by the motivational state of the animals. These experiments make it clear that such error decreases are not merely a function of the degree of "latent learning" which has occurred, but also of the value of the reward which is introduced. In view of this, it is curious that later investigators seeking to verify the Blodgett findings have utilized incentives relevant to but one of the animal's drives. It is perhaps significant that Blodgett's animals, which were both hungry and thirsty, exhibited error decreases upon being rewarded with a wet mash, while Reynolds'

rats, which were merely hungry, did not show any unusual improvement in performance when dry food pellets were introduced (51). A wet mash is probably more eagerly sought after by hungry and thirsty animals than is dry food by merely hungry animals. This interpretation would presumably also apply to the experiment of Meehl and MacCorquodale (48), in which no marked error decreases were observed upon the introduction of dry food.

3. *Prior to the introduction of reward there is present no relevant goal object or no goal object associated with a relevant incentive.* The greater the reduction in time and error scores of the experimental groups *prior* to the introduction of reward, the less it will be possible for these groups to exhibit marked error decreases. In Reynolds' repetition of Blodgett's experiment some of the experimental animals had previously been used in a spatial discrimination experiment involving a simple T-maze. Blodgett (3) has criticized this use of maze-wise animals, arguing that it was unwise to use such animals in view of Anderson's finding (1) that animals which had previously been given a series of rewarded trials on a six-unit T-maze, learned another maze under "non-reward" conditions almost as rapidly as a rewarded control group. Such results may be attributed to the association of certain maze parts, procedures, and the like with a relevant incentive. Maze-wise animals, when trained on an unpracticed multiple-unit maze, may quickly exhibit efficient maze performance even though they are trained without any known reward or under satiated conditions. With such animals it is difficult to create "non-reward" conditions, and hence latent learning is not likely to occur.

II. *Experiments Depending on Exploratory Training*

As in the preceding type of experiment it is primarily the incentive which is manipulated in type II experiments. Instead of being given a series of unrewarded trials in the empty maze, the animal is permitted to explore the maze for an extended period of time. After the exploratory phase of training the animal is either pre-fed in the goal box and then permitted to run one or more test trials in the maze (5, 38, 53), or it is placed directly in the maze and given a series of test trials (9, 15, 28).

Lashley (35) reported an experiment of this type as early as 1918. Many years later, Haney (15) repeated the experiment using, however, a complex 14-unit T-maze and permitting his animals to explore and live in the maze for 18-hour periods on four successive days. A control group was given similar handling and exploration in a simple rectangular alley path. Haney then gave both groups a series of rewarded trials in the maze. The experimental group, which had had previous exploration in the maze, made significantly fewer errors than the control group. Since the time scores of the groups did not differ significantly it is unlikely that a differential delay in reinforcement can account entirely for the observed differences in error scores.

Daub's experiment (9) provides a check on the hypothesis that the operation of doors to prevent retracing may induce some kind of motivation facilitating progress through the maze. Daub gave his experimental and control animals preliminary exploration on a 14-unit T-maze and a rectangular alley path, respectively. Subgroups of both of these groups were then given rewarded trials on the maze either with or without the use of doors. Daub found that his animals demonstrated latent learning and that the use of doors did not make any significant difference in error scores—latent learning occurred both with and without the use of doors.

An experiment using much shorter exploratory and handling periods has been reported by Karn and Porter (28). Different groups of rats were given various types of pre-training experiences, and it was observed in general that the more complete the pre-training experience, the greater the effect upon subsequent performance in the Dashiell maze. Thus, for example, one group of rats, received handling only for several days; another group had both handling and confinement in an enclosure; still another had handling plus exploratory experience in the maze, etc. The results of Karn and Porter suggest, although the differences here are not statistically significant, that unrewarded exploratory experience in the maze has greater effect on maze performance than pre-training which does not include such exploration. These results, together with those of Haney and Daub, argue against the hypothesis that the superior performance of the exploratory animals in these investigations is attributable merely to superior emotional adjustment.

Two experiments by Buxton (5) and Seward (53) employ the method of pre-feeding the animals in the goal box subsequent to exploration and just prior to the single critical test trial on the maze. In Seward's experiment it is fairly certain that the correct choice on the critical trial depended upon the preceding exploration. In one of Seward's control experiments, animals which were treated just like the original group of animals except that they were given no preliminary exploration, failed to take the correct path on the critical trial any more frequently than would be expected by chance.

In a second control experiment, Seward attempted to determine what cues his rats had relied upon, by eliminating differential cues from inside the end boxes for some animals and by eliminating cues other than intra-endbox cues for other animals. The former was accomplished by training rats in a maze in which two identically painted goal boxes were interchanged several times during the four exploration days. To check upon the effect of eliminating differential cues other than those emanating from within the end boxes, Seward gave a second subgroup its pre-feeding experience in one of the endboxes which had been moved from its place in the maze and placed just behind the starting box. Neither of these subgroups was able to respond appropriately on the test trial, and Seward concluded that the animals in the main experiment must have reacted to cues both inside and outside the endboxes. It is possible that the rats in the latter subgroup were learning that food was to be found in the endbox in its *new location.* Seward reports that one rat on its test run tried to climb the back wall of the starting box (it will be recalled that the feeding experience had occurred just behind the starting box for this control group). This control experiment is similar to one of Leeper's (**38**), in which rats were fed and watered in distinctive goal boxes after a preliminary period of exploration on an elevated maze. Before the feeding and drinking stage Leeper removed the food box and water box from the maze to another part of the room. Leeper's results were also negative (although the very rapid trial-and-error learning of these rats suggests some learning during the exploration period). Care was taken in Leeper's experiment to enclose each of the endboxes so that when a rat was in any one of them it could see only the interior of that endbox. Furthermore, the rats "were introduced into the appropriate endboxes individually, being carried in such a manner that it was not possible for them to see the exterior of the boxes into which they were placed" (**38**, p. 25). It is unlikely that an association between eating or drinking and the new spatial locus of the boxes occurred under these conditions. Each of Leeper's three endboxes had distinctive interior textures, but rats are apparently unable to make the required associations when they must rely solely upon such intra-endbox cues during the feeding or drinking experiences. We may summarize these results as follows: If differential cues to which the animals would normally respond during the course of training are removed so that the identifiability of the endboxes is lessened, or if the cues are altered during the course of training so that conflicting expectations arise, rats are unable to solve the problem.

It is possible in any experiment which employs the kind of prefeeding procedure used by Buxton and Seward that the correct response was mediated by a perseverating goal reaction from the endbox in which the animals have just previously been fed. Hull has already illustrated how the performance of Maier's rats might be explained in this way (**25**, p. 236 ff.). A third control experiment by Seward provides a test of this hypothesis. The animals in this experiment, instead of being placed in only one terminal box of the maze in the pre-feeding stage, were placed for two minutes in one endbox and then for two minutes in the other endbox. Food was available as before in only one of the endboxes, and half of the rats were placed in the food box first and half in the empty box first. Thus, if the rats had been responding merely to a perseverating goal reaction which mediated the correct choice, then these control animals should have performed much less efficiently than the animals in the main experiment. This expectation was not confirmed; the percentage of correct choices of the controls (81.3) did not differ significantly from the percentage of correct choices (87.5) of the main group. One of Buxton's control experiments may also be cited as a check upon this interpretation. In this experiment, rats lived in and explored a 12-unit T-maze for several nights just as did the experimental groups. The pre-feeding experience was omitted for these animals, however, and they made their critical runs without finding food anywhere in or near the maze during the trials. No perseverating goal reaction from the eating of food in the endbox could have been operative under these conditions. Although the number of animals meeting the criterion of success on the first test trial was less than the corresponding numbers in the main groups receiving pre-feeding, this number was significantly greater than the number of those successful in a control group whose exploratory training was confined to a straight alley. Thus the efficient maze performance of Buxton's rats apparently depended upon preliminary exploration in the maze and does not seem to be attributable solely to any perseverating goal reaction from eating in the goal box.

Although it is now well established that rats acquire some familiarity or knowledge of a maze by exploring and living in the maze paths, it is still unknown how long the exploratory period must be in order for such learning to occur. The answer to this question will probably depend upon the complexity of the maze path to be negotiated, the number of differential cues reacted to, and many other parameters. Seward gave his animals a total exploratory period of approximately 90 minutes in a single-unit T-maze, while Lashley found that 20 minutes was sufficient to affect performance on his simple two-path maze. Workers using multiple-unit T-mazes have used total exploratory periods of 48 hours or more. Prima facie it would seem reasonable that the longer the exploratory period in the maze, the better should be the learning exhibited on the critical test trials. Since Buxton's animals explored the maze for three, four, six, or nine nights, his experiment provides some data bearing on this hypothesis. Additional amounts of exploratory training did not produce larger percentages of animals that could meet the error criterion. One explanation which Buxton offers for this fact is that individual differences in the rats may have been such that the "per cent successful" simply represented the proportion of each group capable of mastering the task. Apparently Buxton did not give the unsuccessful animals additional trials until each met the criterion of three or less errors; at least, no supporting data are presented. It is possible that all of Buxton's experimental groups received amounts of exploratory training in excess of the optimum for these particular maze conditions, so that no gradient in clarity of knowledge (or degree of learning) was observed with increases in total exploratory time.

Another of Buxton's findings is of considerable interest. His experimental animals which exhibited latent learning also manifested a speed-of-locomotion gradient on the first run through the maze, and a backward order of elimination of blinds. Hull (**23**) made his deduction of goal gradient phenomena on the basis of a conditioning which is progressively weaker as the reactions of the animal are more remote in time or space from the point of reinforcement. Since the animals had never before been given a rewarded run in the maze to the point of reinforcement, the conditioning process on which Hull's deduction is based could not have occurred. Hilgard has mildly criticized Buxton's experiment on the grounds that the results reported need to be supplemented in order to interpret with confidence the gradients of running speed (**21**, p. 285).

To summarize, the problem with which the animal is presented in the foregoing latent learning experiments is primarily one of learning the characteristic features of the environment under non-reward conditions. It has been found that latent learning is most easily demonstrated under the following conditions: The maze environment to be mastered is complex or otherwise yields reliable measurements of individual differences in maze performance; the reward introduced is highly demanded or is associated with a highly demanded goal object; and the response by which the learning is manifested is not associated with a relevant reward during the non-reward trials. It appears, further, that if the means of eliciting an already acquired environmental familiarity is especially emphatic—as with the pre-feeding technique employed by some investigators—the learning may be demonstrated even in relatively simple maze environments. Such learning also seems to require a sufficiency of intra- and extra-maze cues and avoidance of the introduction of contradictory cues.

III. *Experiments Using Weak Irrelevant Motivation*

The alternative of varying the motivational state of the animal in order to alter relationships between drives and incentives, is illustrated in experiments of the third and fourth types (irrelevant-incentive learning). In the former (type III) the animals are trained while satiated for food and water and are presumably motivated during the training period by weak, secondary drives irrelevant to the food and water present in the goal boxes of the maze.

Szymanski's early experiment (62) was the first to utilize this kind of experimental design. Using a modified Hampton Court maze, Szymanski ran several animals when not hungry, for 61 trials through the maze to the home cage in which food had been placed. During this period there was no reliable tendency toward a reduction in time, errors, or in distance traveled per trial. During the critical trials when the rats were made hungry, however, they quickly exhibited efficient maze performance.

Suggestive as these results were, they were not followed up by experiments of similar design until recently, when Spence and Lippitt (59) initiated the Iowa series of latent learning experiments. Briefly, the Spence and Lippitt experiment demonstrated that rats which had been satiated for food and water and run on a single-unit Y-maze with food in one end of the Y and water in the other, learned the location of the goal objects. On the critical test trials, the animals which exhibited preferences for the "water side" during training were made hungry, while those showing a "food-side" preference were made thirsty. The animals made significantly correct choices on their first test run. However, on the second test run the next day, when their need was shifted, they did not make a significant shift to the correct side (R. Lippitt, personal communication).

Meehl and MacCorquodale (47) repeated the above experiment with some slight modifications in design. A single-unit T-maze was used; the terminal sections of the maze were so placed that a rat made a right turn at the end of the right alley and a left turn at the end of the left alley. During the 40 training trials under satiation, water was in one terminal section of the maze and food in the other. During the initial test trials the rats, now hungry or thirsty, responded appropriately. Although performance deteriorated (as in the Spence and Lippitt experiment) when the drive was shifted on subsequent trials, the proportion of responses to the appropriate side on all five test trials was significantly greater than chance.

The satiated group of animals in Kendler's experiment (30) received experimental training similar to the latent learning groups in the above experiments. Although this group of rats when made hungry or thirsty consistently chose the appropriate alleys more frequently than 50 per cent, Kendler did not attempt to draw any conclusions as to whether they exhibited latent learning. Since no information concerning the pre-test performance of these groups is reported, it is impossible to evaluate this finding.

On the theoretical level, perhaps the most crucial problem is that of determining whether certain alleged secondary reinforcing agents can reasonably be interpreted as mediating irrelevant-incentive learning. This problem has been attacked by investigating the following questions: (1) *Does the irrelevant incentive itself serve as a secondary reinforcing agent?* and (2) *Is the degree of irrelevant-incentive learning exhibited a function of the delay in reinforcement of the critical S–R sequence?*

1. The evidence relating to the first question is, with one exception, negative. This is indicated by the fact that, other things being equal, rats satiated for food do not select a food alley in preference to a non-food alley. This has been demonstrated both for satiated rats (**63**) and for thirsty rats (**13**, **61**, **72**).[4] The exception itself is instructive from a methodological standpoint. MacCorquodale and Meehl (**43**) ran animals satiated for food and water in a single-unit T-maze with one goal box containing food and water and the other empty. Half of the animals found food and water on the left; the other half found it on the right. Although an attempt was made to match the animals in the strength and direction of position preferences, on the *initial* training day the group finding food and water on the left made 20 per cent more left turns than the group finding food and water on the right. This initial inequality before the animals had experienced food in the maze, detracts somewhat from the finding that there was a statistically significant difference between the two groups in the total number of left turns, with the group finding food and water on the left making the greater number of left turns. Further, these investigators employed feeding procedures which did not establish any definite, known feeding rhythm. Also, they attempted to satiate their rats for food and water simultaneously and employed a dry, "stick"-type of food. Although MacCorquodale and Meehl report that no rat ever ate or drank in the maze during the training trials, one might expect that, with a detention period longer than 30 seconds, some animals treated in this way would exhibit consummatory responses toward the incentives. A more rigorous test of the hypothesis in question would involve procedures which clearly rule out inadequate satiation due to the prior ingestion of dry food (perhaps causing thirst) or to the unknown periodicity of feeding habits. In view of these facts, and in consideration of the negative evidence referred to above, it seems reasonable to conclude that the irrelevant incentive in those experiments in which motivation is carefully controlled does not serve as a source of primary or secondary reinforcement.

2. A method of testing the second of the above questions has been suggested by Meehl and MacCorquodale (**47**, p. 388). If reinforcement arises from the procedures of removing the rats from the terminal sections of the maze, returning them to the home cage, etc., then the greater the distance between the irrelevant incentive (e.g., food) and this point of alleged reinforcement, the poorer should be the irrelevant-incentive learning manifested. This would follow if it is assumed that such learning is dependent upon the strengthening of connections between fractional consummatory responses (elicited by the sight of food) and the distinctive proprioceptive stimuli associated with the

---

[4] In an unpublished experiment (type IV) by the writer, involving approximately one hundred animals, it was also found that thirsty rats satiated for food did not tend to select a food pathway in preference to a non-food pathway.

turning responses in the terminal sections of the maze. Thistlethwaite (63) found, however, that changing the spatial location of the food did not have any particular effect upon the irrelevant-incentive learning of satiated rats, nor was the degree of such learning a function of the number of alleged reinforcements of the critical S–R sequence.

Attempts to study the relative potency of "cognition" and reinforced trials (47), or to compare the amount of learning occurring with or without (relevant) motivation (30), may be seen as different ways of obtaining data on a single question: *Does learning the location of, and means of arriving at, a given incentive occur more readily in the presence of reinforcement, or is such learning equally efficient in the absence of reinforcement?* Here the answer will depend upon a variety of parameter values. One of these appears to be the complexity of the maze path to be learned. Thus Tolman and Honzik (70), using a 14-unit T-maze, found that a group of rats rewarded only from the eleventh trial on made significantly fewer errors on subsequent trials than a control group which had been regularly rewarded from the beginning. The authors point out that the absence of a food reward during the early trials meant that the experimental group explored the maze much more thoroughly, so that when this group was later rewarded its superior familiarity with the maze permitted more efficient maze performance. Kendler, however, found that his rats learned the location of incentives in a simple T-maze more readily in the presence of relevant drives. Tolman and Honzik manipulated the incentives, whereas Kendler varied the motivational state of his animals. If the interest is chiefly in comparing learning by "reinforcement" with that by "cognition," then varying the incentive relationships is preferable to the technique used by Kendler, since the latter introduces drastic motivational differences between satiated and deprived animals and involves the Hullian assumption that food and water would allegedly act as secondary reinforcers for the satiated animals. Another parameter which must be taken into account in any critical discussion of the above question is the strength of the *irrelevant* drives operative. There is some evidence (59) that irrelevant-incentive learning occurs *less* readily when there are strong irrelevant drives present than when this is not the case. The presence of a strong thirst drive in one of the Spence and Lippitt experiments (60) to be discussed subsequently apparently had the effect of hindering learning of the food location.

IV. *Experiments Using Strong Irrelevant Motivation*

Finally, there is to be considered the variety of irrelevant-incentive experiments in which rats are trained under a strong drive. Upon the completion of training, the animals' motivation is reversed and the test trials are given under a different drive. If thirst was the dominant motivation during training, hunger is usually the drive during the critical test runs (or, occasionally, hunger for some test trials and thirst for others). Any evidence of an appropriate modification of behavior or of a significant increase in the number of choices of that alley of the maze which contains the goal object for which the rats are now motivated, indicates irrelevant-incentive learning.

The discrepancies between various experiments of this type have not as yet been adequately explained. A critical review of these experiments casts serious doubt upon the possibility of isolating any necessary and sufficient condition for irrelevant-incentive learning under strong motivation. The reasons for this doubt will, it is hoped, become clear as we proceed. In the following sections are evaluated various experimental and theoretical attempts to isolate the crucial variables determining whether or not such learning occurs in this domain.

1. *Strength of irrelevant motivation.* In reporting the facts that irrelevant-incentive learning occurred under satiated training but did not occur when training was under a strong thirst drive, Spence and Lippitt (59) offered the tentative generalization that "latent learning does not occur in the situation where the animals 'perceived' the subsequent test goal object (food) while motivated for another goal object (water), but latent learning does occur where complete satiation made for no particular goal directedness in the experimental setup during the training series." Within a short time, however, Deisenroth and Spence (10) reported in abstract a positive case of irrelevant-incentive learning with training under thirst (of unspecified degree), and many other instances of irrelevant-incentive learning in the presence of strong motivation have since been reported (61, 68, 72).[5] The above conclusion must therefore be rejected, although it might still be maintained that such learning occurs more readily when the existing irrelevant drives are not too strong. Since experimenters working in this area have, apparently without exception, used deprivation periods ranging from 18 to 24 hours, there are no clear supporting data. However, the generalization that the stronger the drive the poorer the irrelevant-incentive learning, would seem to be a reasonable inference from the Spence and Lippitt experiments.

2. *Nature of irrelevant motivation.* The experiment of Hayes (17) lends some support to the view that such learning occurs more readily when training is carried out under hunger than when the rats are

---

[5] The findings of Leeper (38) have been neglected by many investigators. Priority should be assigned to Leeper, who reported an experiment suggesting positive irrelevant-incentive learning of this type as early as 1935.

trained under thirst. Hayes found that rats trained under 18 hours' water deprivation failed to respond appropriately to food in a straightaway when tested hungry, whereas rats trained under 23 hours' food deprivation responded appropriately to water when made thirsty. However, results obtained by other investigators who trained some or all of their animals under the hunger drive are at best only partially consistent with this finding. Kendler (**31**) found that neither the animals trained thirsty nor those trained hungry in his experiment gave evidence of irrelevant-incentive learning. The response of selecting the correct alley of a T-maze, which was required in this experiment, is qualitatively different, however, from the response of merely stopping midway down the path of a straightaway as in the Hayes experiment. Christie obtained strikingly appropriate responses on both hunger and thirst test trials with rats which had been trained under hunger motivation. However, since Christie's results were obtained with animals which had been given a variety of pre-training experiences, it is uncertain whether these results are in any way attributable to the fact that training was under hunger.

3. *Strength of preferential tendencies.* Walker, Knotter, and De Valois (**72**) have called attention to the relation between the strength of position preferences and the degree of latent learning manifested. They found that over 50 per cent of their "weak"-preference animals modified their performance appropriately when the motivation was changed from thirst to hunger, whereas only about 5 per cent of the "strong"-preference animals modified their performance. It is significant that in many of the experiments yielding negative results (**31**, **60**, **71**) conditions were such that the animals acquired pronounced turning preferences during the course of training. In other experiments yielding negative results, where asymmetry of reward did not cause a development of preferences, the authors unfortunately do not report what proportion of their animals exhibited strong preferences. It is therefore impossible to say to what extent the failure to find irrelevant-incentive learning in these experiments (**14**, **33**) was attributable to overly-strong preferential tendencies. That this factor is a basic one, however, is indicated by the facts that asymmetry of reward and asymmetry of intra-maze cues appear to be associated both with the development of strong preferences and with a failure to manifest irrelevant-incentive learning. We shall now consider each of these features of experimental design in greater detail.

4. *Asymmetry of reward.* The usual experiment of this type involves a single-unit T- or Y-maze, one alley of which leads to food and the other to water. Definite preferential tendencies are acquired during the training trials because the rats which are thirsty (or hungry) quickly learn to select on their free trials that alley which leads to the appropriate incentive. One method of avoiding this outcome is to place the re-

ward which is appropriate during training in *both* endboxes (the irrelevant incentive being placed in only one alley of the maze). Experiments in which there was asymmetry of reward have generally yielded negative results (31, 60, 71). On the other hand, in one of the experiments in which positive results were obtained (72), the reward (water) was placed symmetrically with respect to the choice point. However, this condition (symmetry of reward) does not seem to be necessary for irrelevant-incentive learning in view of the fact that others (8, 10) obtained positive results without it. And it cannot be a sufficient condition in view of the negative results obtained when the reward was placed on both sides of the maze (14, 33). Further, Gleitman (13) reports that his thirsty rats did not learn the food location when no water was available during training on either side of a T-maze (a condition which might be called "symmetry of non-reward"), although these results are contradictory to those of Strange (61). From one point of view symmetry of reward might make for better learning by lessening the likelihood of strong position preferences. On the other hand, one might suppose that it would make for poorer learning by lessening the identifiability of the paths. It is probably more difficult for a rat to make a functional distinction between a water alley and a water-and-food alley than between a water alley and a food alley. In order to demonstrate this, future experiments might be carried out which would compare the irrelevant-incentive learning of animals trained satiated and run on both types of maze arrangements.

5. *Asymmetry of intra-maze cues.* One might expect that the more distinctive the cues associated with the two alternative choices, the more readily animals will learn "what leads to what." Kendler's experiment (31) sought to test this hypothesis. Kendler used a maze in which the left half of the stem and all doors, walls, and floor on the left were painted black, whereas the right half of the stem and the right alley were left unpainted. In addition there was a black curtain in the entrance to the left alley and a white curtain in the right alley. Yet no irrelevant-incentive learning was observed in the performance of Kendler's ten animals. Kendler and Mencher (33) used similar sets of differential cues and also failed to get irrelevant-incentive learning. However, it has been repeatedly observed (33, 47) that rats tend to prefer black alleys and curtains over those which are white.[6] This observed concomitant variation of preferential tendencies and differential intra-maze cues suggests the difficulty of adequately assessing the above hypothesis. The most drastic differences in intra-maze cues may have to be deliberately avoided in order to tease out the effects of dis-

---

[6] In a recent repetition of the Kendler and Mencher experiment, Rollins found that *every one* of his 36 rats exhibited a preference for the alley painted black (R. A. Rollins, personal communication).

tinctive, but non-disturbing, cues. What is needed, it seems, is a sampling of sets of intra-maze cues sufficiently distinctive to be differentiated easily but not so different in appeal that one set is consistently selected over the others. It is perhaps significant that Christie's (8) most appropriately responding groups were run on a maze painted a flat gray throughout, affording only differential light and spatial cues.

6. *Forced trials.* The experiment by Walker, Knotter, and DeValois (72) was designed to check upon the possibility that irrelevant-incentive learning may not occur as readily when some trials are forced. These investigators ran three groups of animals in a Y-maze with both of the terminal boxes containing water and with intermediary boxes (one of which contained food) between the choice point and endbox on each side. One group was permitted to choose freely on all 40 training trials. A second group was given 50 per cent forced trials, and a third group was given 100 per cent forced trials. A significant tendency for the animals to go to food when made hungry was demonstrated in the group receiving all free trials and in the one having 50 per cent forced trials. There did not appear to be any appreciable difference in the degree of irrelevant-incentive learning exhibited by these two groups. On the other hand, the modification in the performance of the group with 100 per cent forced trials, while in the appropriate direction, was small and did not differ significantly from chance. The effect of forcing the animal down certain paths on some of the training trials has not yet received sufficient experimental study to warrant any general conclusion. Carr (6) has summarized the results of a variety of experiments dealing with the influence of guidance, or forcing, on the learning of complex, multiple-unit mazes. As might be expected, it has been found that some mechanical guidance aids the learning of complex maze paths, and that the number of forced trials and the stage of learning at which these trials are administered are important variables determining the effectiveness of such guidance. Whether these conclusions are applicable to the problem of forced trials in the usual irrelevant-incentive experiment is uncertain. Carr and Koch (7) found that rats run on a single-unit T-maze in an alternation problem made consistently better records when all 800 trials were free then when 80 per cent of the trials were forced. It might be expected that the effects produced by forcing an animal to take the alley of a simple T-maze leading to an irrelevant incentive would be quite different according to the preferences of the animal, to the strength of the dominant drive, and to the probability of finding appropriate reward. Some investigators of irrelevant-incentive learning have reported that forcing to the non-preferred side in simple two-path mazes produced definite disturbances in their rats (**31**, **71**).

7. *Discriminating perception of the irrelevant incentive.* Leeper has suggested (40) that certain irrelevant-incentive experiments yielded negative results because the experimental conditions were not favorable

for the perception of the undesired incentive. One of Leeper's earlier experiments (38) closely paralleled the Spence and Lippitt experiment, since during the first four or five days the animals were run hungry in a maze with both relevant and irrelevant incentives. It should be noted however, that Leeper's rats were trained hungry during this initial period, whereas those of Spence and Lippitt were trained thirsty. On the first trial under thirst, 13 of Leeper's 23 rats went to the now-correct side, and on the first trial of the second thirsty day, 19 of the 23 went to the now-correct side. This performance was far superior to that of the Spence and Lippitt (60) and Kendler (31) animals. Leeper observed that in his experiment, "on a wrong run, the situation was such as to guarantee that, immediately after making the run, the rat would approach the pan containing the undesired goal material and perceive discriminatingly what it contained" (40, p. 105). Leeper's interpretation of the negative irrelevant-incentive experiments has been vigorously criticized (33, 58). It is argued that the Kendler and Mencher experiment which required the animals to select water from one of five glasses (some of which contained food) satisfied this condition, and the animals nonetheless failed to learn the location of food. It is doubtful, however, whether this experiment constitutes a convincing refutation of Leeper's hypothesis. Kendler and Mencher used 50 per cent forced trials and employed a maze with clearly differentiated white and black alleys, and report that their animals exhibited a definite preference for the side of the maze painted black. Whether the position preferences acquired under these conditions were so strong that irrelevant-incentive learning was precluded cannot be determined from the results reported. Furthermore, the technique used by Kendler and Mencher of comparing *over the total 20 test trials* the performance of a group which found food on the same side as during the original thirst training with that of a group which found the food reversed to the opposite side, presupposes that learning is a gradual process of strengthening habits. On the basis of such non-continuity interpretations of learning as have been defended by Krech (34) and Lashley and Wade (37), one might predict that the reversed group would rather quickly give up a maladaptive hypothesis and learn within a very few trials to go to the now-correct side. The comparison of total performance scores over a long relearning period might well obscure, rather than enhance, any initial superiority of the constant group.

On the other hand, a comparison of Hull's results (24), obtained from an experiment in which rats were alternately hungry and thirsty on successive days, and Leeper's results, obtained from a similar experiment,[7] strongly suggests that it is a matter of considerable importance whether the animals on their wrong trials simply find a closed door at the end of the path or whether they find the incentive for which they are satiated. If Hull's curve (24, p. 263) is smoothed one might come to the conclusion that his rats reached a 70 per cent accuracy on first trials by the end of 120 days (15 eight-day periods of training). In Hull's procedure the rats were not permitted to find the irrelevant incentive on their incorrect runs. On the other hand, Leeper permitted his rats to find the irrelevant incentive on the wrong runs and found that they learned the same problem to the same level of accuracy in about 24 days (this period was estimated by smoothing Leeper's first-trial curve). Leeper concluded from his experiments that the animals "must learn almost as much from their incorrect responses as they do from their correct responses" (38, p. 37). Apparently irrelevant-incentive learning made the difference in the outcomes of these two studies. Inasmuch as an experimental setup requiring the animals to perceive the irrelevant object discriminatingly made such a vast difference in learning this problem, it seems very probable that such a setup would facilitate irrelevant-incentive learning. The Kendler and Mencher experiment and others, however, argue against its sufficiency.

---

[7] The comparison discussed here is between the performance of Group III-C in Leeper's experiment and that of Hull's main group. This comparison is more easily interpreted than others which might be made, because both of these groups were run in similar mazes which differed only with respect to the construction of the endbox or endboxes.

8. *Discrimination of internal stimuli.* The problem which the animal is required to solve in this type of irrelevant-incentive experiment is very similar to the problem which Hull (24) gave his animals in the experiment just discussed. In both setups the problem for the rat is essentially one of learning to take one path of a maze if it is thirsty and the other path if it is hungry. There are, of course, important differences between the two problems, namely: In the Hull experiment hunger and thirst drives were alternated on successive days, and the correction method was employed so that the animals always received reinforcement on a given trial. The extremely slow learning of Hull's rats led him to suggest that there is a lack of distinctness between the hunger and thirst drives of the rat (24, p. 270). In a later formulation, Hull maintains that the problem is essentially one of learning to discriminate different patterns or combinations of stimuli (27). These stimuli are said to include the drive stimuli peculiar to hunger and thirst, distinctive proprioceptive stimuli arising from postural movements, etc. More recently, Kendler (29) has reported an experiment which casts doubt upon Hull's explanation in terms of two stimulus patterns each conditioned to a specific response. Heron (20) has also questioned, on the basis of experimental data obtained from a closely parallel experiment, whether the animals were actually making the differentiation upon the basis of drive stimuli. Heron's results suggests that his rats were using stimuli from a full versus a partially empty stomach.

Regardless of what mechanisms underlie the differentiation, however, there remains the possibility that whatever stimuli, or patterns of stimuli, are used are not sufficiently distinctive to become quickly and clearly associated with the routes to the appropriate incentives.[8] If this is the case, it is reasonable to suppose that pre-training in making discriminations between hunger and thirst states and associating each with a distinctive response would aid the animals in making the modifications of performance required in this type of irrelevant-incentive experiment. In an unpublished study, Lippitt found that preliminary "need differentiation" exercise facilitated the differential motivational control of responses on the test trials. Lippitt's animals were satiated for food and water during training in the irrelevant-incentive experiment proper. Although the number of his animals was small, previous exercise in the discrimination of drive states and the association of each with going to one end of a straightaway, facilitated appropriate responses on the test trials, particularly on later test trials when the motivation was reversed. More recently, in an unpublished experiment by the writer, similar preliminary practice definitely facilitated learning when the training series was given under strong irrelevant motivation (thirst). Even if the conclusion is accepted that such preliminary training is efficacious, much more work will need to be done to determine how it mediates better irrelevant-incentive learning.

9. *Previous experience.* We have already observed that previous exploratory experience in the test maze itself represents one major condition under which latent learning occurs (type II experiments). Does previous exploratory exercise in non-maze environments aid subsequent performance in the maze? Several experiments suggest that it does. Hebb (18), for example, compared the problem solving of animals reared to maturity in cages with others which were allowed frequent exploration of a varied environment. The latter were reared at home as pets with much of their time spent outside their cages. Hebb reported that the group with the greater infant experience exhibited a definite superiority in various tests of problem solving.

Christie's provocative series of experiments (8) was designed to test a number of hypotheses concerning the effect of early experience in the irrelevant-incentive learning of rats. Christie subjected some of his animals to a complex regimen of water and food deprivations from the time of weaning until three months of age (for approximately 60 days). When these animals were later trained under conditions similar to Kendler's (31), Christie found that these early deprived groups exhibited irrelevant-incentive learning. One of the groups which was deprived from infancy was permitted to satisfy its needs after the deprivation periods only by exploring an elevated path and table-top on which food and water were placed. The animals in this group remained there for 24 hours after the termination of each deprivation period. Since food and water were presented in close proximity to each other on the table-top, the rats were not required to associate distinctive turning responses with particular drive states. A second group which had food and water in its cages at all times was given the same exploratory exercise as the first group, i.e., they were put on the pathway leading to the table and allowed to remain there for 24 hours at a stretch. The position of the pathway was varied on successive occasions so that in no case did the animal find the path to food and water occupying the same spatial configuration on any two successive days. Christie found that both groups when later trained hungry and tested on a T-maze for irrelevant-incentive learning went to water with high degrees of accuracy on the two test trials under thirst. Moreover, the exploratory group without any previous deprivation experience was superior to the deprived exploratory group on the two hungry test runs.

---

[8] Heron's technique of satiating the rats for a given incentive two hours before the day's trials probably produced more distinctive internal stimuli (i.e., more completely full versus empty stomachs) than the technique used by others of having the incentive available in the cages continually.

In line with this are the experimental findings which Bruce (4) has referred to as "perceptual learning." Bruce gave several groups of rats ten days' practice in drinking in one square of a field which was later the goal-square, and observed that all three groups of rats that had had the practice learned more rapidly than did animals that did not have this practice. The problem which these rats had to master was to run from one end of the open field to the other end where water was available. Bruce writes regarding the animals which received previous drinking practice, "if we analyze this behavior . . . it seems possible that perceptually the animals have largely solved the problem before they have actually moved from one end of the field to the other" (p. 171).

The foregoing experiments, together with the studies of Harlow (16), have recently made an issue of the avoidance of maze-wise or task-wise subjects in learning experimentation. If it is true that the behavior of the organism is limited by the hypotheses which it brings to the task as a result of previous experience, then it must be granted that principles of learning derived only from experimentally naive subjects will be severely truncated and so much the less applicable to ordinary human or social learning. The crucial question to be asked of any latent learning or irrelevant-incentive experiment involving subjects with previous experience is whether the experience was such that the *response* which is finally taken as evidence of such learning could have been rewarded, or associated with reward, as a result of this previous experience. In Reynolds' experiment (51), which has been criticized for the use of maze-wise animals, it is not clear whether the response of running the six-unit T-maze could have been associated with reward as a result of the previous maze experience. On the other hand, in the above experiments the correct response on the later test trials definitely does not appear to have been mediated by any peripheral responses which became strengthened as a result of the previous training.

10. *Amount of training.* Spence and Lippitt (60) have pointed out that their thirsty rats learned to go with almost 100 per cent accuracy to the water alley of a Y-maze after having had an average of only seven to eight experiences with this goal object. They have argued that *if* thirsty rats (satiated for food) can learn which path leads to food without reinforcement, then their rats should have done so after having had the number of experiences of food which their experiment provided (an average of 26 trials to the food alley). Similarly Walker, Knotter, and DeValois (72) have inferred that a response acquired under an irrelevant drive is much weaker than one acquired under a relevant drive. This conclusion is based on the observation that a training series of 40 trials under thirst motivation was adequate for rats to acquire a water-going response overcoming even the strongest position preferences (71), whereas the food-going response demonstrated when rats were made hungry after a similar period of training was not strong enough to overcome the strongest position preferences (72). Such comparisons, however, only indirectly bear upon the question of the relative efficiency of learning with reinforcement as compared with learning in the absence of reinforcement. The problem which the rat must master in this type of experiment is not simply that of going to one end of a Y-maze for food. As was pointed out above, the problem is essentially one of learning to take one path of the maze when thirsty and the other path when hungry. Hull's experiment (24), checked and supported by Leeper's experiment III-A (38), provides us with relevant data as to how rapidly rats can master this problem when every trial leads only to reinforcement and not to perception of the irrelevant incentive on the wrong runs. Smoothing Hull's curve, we may estimate that his rats when run under the correction method required approximately six hundred trials to attain a 70 per cent level of accuracy. As we have already seen, under the more favorable conditions in Leeper's experiment III-C, rats required approximately 24 days of training, or about 120 trials, to reach and maintain a 70 per cent level of accuracy on first trials (based on a smoothing of Leeper's curve).

It is true that most irrelevant incentive experiments have employed the non-correction method and that some of them have been designed to facilitate perception of the irrelevant incentive on the wrong runs. However, one might expect that investigators seeking to determine whether rats run under a strong irrelevant drive can learn to master this problem without reinforcement would give the animals, at least, a similarly large number of training trials. Actually we find that the number of training trials in the relevant studies have ranged from 12 to 60. It should be obvious that no irrelevant-incentive experiment involving a training period falling in this range can provide a convincing demonstration that rats, run under a strong irrelevant drive, typically do not learn the problem in the absence of reinforcement. On the other hand, the fact that some investigators (8, 72) have obtained positive results with relatively short training periods indicates that when conditions are especially favorable smaller amounts of training may suffice. It is possible that previous exploratory exercise, or preliminary training in the discrimination of drive states, serves as a substitute for otherwise inadequate amounts of training. The Leeper and Hull experiments, it should be pointed out, were so designed that both pathways were "correct" an approximately equal number of times during the training series. On the basis of these experiments we can infer that longer periods of training under conditions which do *not* produce any strong asymmetrical tendencies will probably facilitate irrelevant-incentive learning.

Summarizing, it appears that the problem with which the animal is presented in the foregoing studies of irrelevant-incentive learning is essentially one of learning relationships between characteristic features of the environment and various drive states. Irrelevant-incentive learning seems to occur most readily under the following conditions: weak irrelevant motivation, hunger rather than thirst as the irrelevant drive, weak position preferences, symmetry of reward during training, free rather than forced trials, discriminating contact with the undesired goal object, previous exploratory or "need" differentiation exercise, and large amounts of training under conditions which do not produce strong asymmetrical preferences.

Theoretical Interpretations

Some of the above experimental studies provide a rather conclusive answer to the first question raised in this survey: Is reinforcement a necessary condition for learning? Since the sources of reinforcement are pervasive and often subtle in nature, a careful examination of the evidence and its relationship to such theories is necessary. Unfortunately, limitations of space forbid anything more than the barest outlines of the various deductions which might be made from reinforcement theories of learning.

If the indispensability of reinforcement for learning is to be granted, it must be possible to demonstrate for *each* instance of latent learning or of irrelevant-incentive learning (1) that some source of reinforcement was operative in the experimental setup and (2) that the changes in responses which are taken as evidence of the learning can be deduced on the basis of this alleged reinforcement. Much of the equivocality of "crucial" tests in this area is attributable to the failure of theorists to state the conditions under which certain alleged sources of reinforcement are present and active. In the strict sense, as long as such statements are lacking, reinforcement interpretations of latent learning or irrelevant-incentive learning are unverifiable. However, if the second of the above requirements can be satisfied, then the set of reinforcement postulates in question may have some degree of plausibility, this unverifiability notwithstanding. The evidence, however, is of such a nature that it appears to be impossible to deduce, from any existing reinforcement theory, the behavioral changes indicating latent learning or irrelevant-incentive learning in each of the reported experiments. One intractable instance alone would suffice to answer our question. Actually, there are many examples of such learning which are difficult, if not impossible, to deduce from present reinforcement postulates.

1. Several instances (**2**, **19**, **70**, **73**) of type I experiments are difficult to explain in reinforcement terms. Possible sources of reinforcement in the Blodgett variety of latent learning, which immediately come to mind, are satisfactions of curiosity and escape from the maze. Conjectures based upon such alleged reinforcements have already been critically evaluated by Leeper (**39**, pp. 27–28) and by Hilgard (**21**, p. 284), hence we need not repeat these considerations here. The chief shortcoming of such conjectures is that the abrupt behavioral changes following the introduction of reward simply have not been shown to be deducible from the reinforcement postulate. More recently, Meehl (**46**) has suggested that such abrupt behavioral changes might be deduced by means of an additional assumption regarding the asymptotes of the habit growth functions involved. Meehl writes:

> A neo-Hullian . . . might assume that when a response undergoes an increment in strength on the basis of a minimally reinforcing agent (that is, one in which the asymptote of the acquisition of habit strength is relatively low), then, if subsequently a strong reinforcement is introduced, the parameter in the new growth function which determines the rate of approach to the new asymptote is greater than it would have been without the original learning. Since in the Blodgett design there is evidence of acquisition of differential habit strengths during the latent phase, such a postulate would lead to a preservation of the Strong Law of Effect (**46**, pp. 70–71).

There are considerations which make this explanation implausible. First, it is assumed that there must be an acquisition of differential habit strengths during the latent period, or that the finally-to-be-strengthened receptor-effector connections are being strengthened during the latent period. However, the experiment of Herb (**19**) demonstrated that rats when properly motivated could perform responses *inimical* to those which were, by this interpretation, slightly rewarded. It will be recalled that Herb's rats exhibited an ability to enter the culs when food was placed there, even though the effect of the learning which presumably occurred under the "minimally reinforcing" conditions should have been the formation of connections so that stimuli from the maze should have evoked reactions equivalent to more or less entering upon and running the true path. Similarly, the experiment of Wallace, Blackwell, and Jenkins (73) showed that latent learning occurred even when the rats were removed at chance points in the maze and even when the goal box was absent so that the crucial response could not have been practiced and reinforced. This procedure seems to rule out clearly the differential reinforcement of the finally-to-be-rewarded response.

2. In those instances (5, 9, 15, 53) of latent learning which depend upon previous exploratory training (type II), satisfaction of curiosity and escape from the maze again come to mind as possible sources of reinforcement. Buxton (5, pp. 4–7) has critically evaluated the former possibility, and Leeper's criticism (39) is again applicable. In short, the difficulty here is that the alleged exploration-reinforcement occurs in the presence of "incorrect" as well as "correct" connections and offers no basis for the differential conditioning of correct responses. In the Daub and Haney experiments, the rats were presumably removed from the maze at chance points. Buxton's rats were introduced into, and removed from, the maze at various positions which were systematically varied so that no parts of the maze were favored or handicapped, in learning, by this treatment. Similarly, Seward's procedure of removing his rats from a simple T-maze does not appear to involve any differential strengthening of right or left turns. Escape-reinforcement, therefore, does not appear to provide the means by which such learning can be deduced. A third type of explanation is possible for those experiments in which rats are fed in the goal box immediately prior to the critical test runs (5, 53). It might be argued that the correct response in such experiments is "essentially a case of the perseveration of a goal reaction . . . from one stimulus situation to a closely following one . . . " (25, p. 243). However, the control experiments by both Buxton and Seward demonstrate rather clearly that their results cannot be entirely accounted for on such a basis. Finally, the application of Hull's concept of anticipatory goal reactions, in a manner analogous to his analysis of Maier's four-pathway reasoning experiment (25, p. 221 ff.), does not appear to be possible here because all of the habit segments involved in the finally correct response are not securely anchored to any common pure-stimulus or symbolic act. On the other hand, Seward's theoretical derivation (52, 53) of latent learning in such types of experiments demonstrates that such learning, while not readily reducible to reinforcement terms, may be interpreted with some adequacy in S–R terms. However, Seward's construct of the "surrogate response" is essentially central in character; hence his analysis is not to be construed as one in terms of the association between peripheral stimuli and reactions.

3. Positive evidences of irrelevant-incentive learning in type III

experiments (47, 59, 63) similarly require explanation. Meehl and MacCorquodale have attempted to provide a reinforcement interpretation of the learning exhibited by their satiated animals. Their derivation assumes distinctive turning responses (at each end of the maze arms) which become conditioned to fractional consummatory responses elicited by the sight, smell, and touch of the irrelevant incentives. Hull (25) and Miller (49) have similarly stressed that special devices calculated to induce very distinctive postures and goal reactions should be introduced before fractional anticipatory goal reactions can be expected to play a directive role in the utilization of habits. Meehl and MacCorquodale have suggested that one method of testing their assumption would be to reverse the turns required at the ends of the maze arms. They have predicted that the stimuli associated with the "wrong" turning responses would then become associated with fractional eating and drinking responses, and the animals, when made hungry or thirsty, would fail to go to the appropriate alley and very probably go to the inappropriate alley. Maltzman (45) has presented a closely analogous derivation which relies upon the conditioning of fractional anticipatory responses and hypothetical proprioceptive stimuli associated with choice-point turns.

As yet the only part of the above derivations to receive experimental study has been the set of assumptions regarding the relationships between habit strength and various aspects of the alleged reinforcements. Since irrelevant-incentive learning does not appear to be related to any *demonstrable* secondary reinforcing effect from the sight, smell, and touch of food, nor to the degree of delay in escape or social reinforcement, nor to the number of alleged reinforcements of the critical S–R sequence (63), it may tentatively be concluded that such learning is not dependent upon reinforcement.

4. Lastly, certain instances of irrelevant-incentive learning (8, 61, 72) in type IV experiments may be considered for their compatibility with reinforcement principles. A reinforcement derivation similar to the foregoing derivations has been presented by Spence.[9] For the situation in which food is in one goal box of the T-maze and water is in the other, the derivation would presumably assume that rats trained thirsty obtain primary reinforcement in the water alley and secondary reinforcement in the food alley. Elsewhere, Spence has stated that the sight and smell of food will have the capacity to elicit the reinforcing mechanism (55) and that such secondary reinforcement regularly plays an important role in maze learning (57). The stimuli in the food

---

[9] Unpublished paper reported at the 1941 meeting of the Midwestern Psychological Association; results presented in abstract (10). Kendler (29, 32) has reported summaries of Spence's derivation. Several previous papers provide the basis for this use of the construct of the fractional anticipatory goal response, especially those of Hull (22, 25, 26) and Miller (49).

2. Such unreinforced learning has been exhibited in a variety of experimental situations and confirmed by studies sampling different conditions within each of several differentiated classes. Certain conditions are favorable to such learning, and other conditions are unfavorable. These conditions have been tentatively identified within certain broad limits: some appear to be predominantly motivational and perceptual in character; others have to do with controlling the development of preferential tendencies; and still other conditions seem to have only a psychometric or mensurational significance.

Much of the difficulty of interpreting the irrelevant-incentive experiments reported to date seems to be directly traceable to the fact that no single investigator can adequately sample a variety of experimental conditions, while uncoordinated work by different investigators seemingly leads to the collection of many sets of data which are not comparable and which therefore can be assessed only with considerable uncertainty. Since the study of latent learning and irrelevant-incentive learning has always been, at least implicitly, a study of the motivational and perceptual conditions which affect learning, it may be expected that many of the questions raised by this series of studies will occupy an important role in future learning experimentation. It is proposed that future research in this area be accompanied by a greater degree of cooperative planning among the interested investigators, using if possible mutually acceptable or standard experimental designs.

BIBLIOGRAPHY

1. ANDERSON, E. E. The externalization of drive: III. Maze learning by nonrewarded and satiated rats. *J. genet. Psychol.*, 1941, **59**, 397–426.
2. BLODGETT, H. C. The effect of the introduction of reward upon the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1929, **4**, 113–134.
3. BLODGETT, H. C. Reynolds' repetition of Blodgett's experiment on latent learning. *J. exp. Psychol.*, 1946, **36**, 184–186.
4. BRUCE, R. H. The effect of varying the place of the fractional anticipatory-consummatory response upon the rate of acquiring a simple learning problem. *J. genet. Psychol.*, 1943, **63**, 165–175.
5. BUXTON, C. E. Latent learning and the goal gradient hypothesis. *Contr. psychol. Theor.*, 1940, **2**, No. 2.
6. CARR, H. A. Teaching and learning. *J. genet. Psychol.*, 1930, **37**, 189–219.
7. CARR, H. A., & KOCH, HELEN L. The influence of extraneous controls in the learning process. *Psychol. Rev.*, 1919, **26**, 287–293.
8. CHRISTIE, R. The effect of some early experiences in the latent learning of adult rats. (In preparation.)
9. DAUB, C. T. The effect of doors on latent learning. *J. comp. Psychol.*, 1933, **15**, 49–58.
10. DEISENROTH, C. F., & SPENCE, K. W. An investigation of latent learning in the white rat. *Psychol. Bull.*, 1941, **38**, 706. (Abstract.)
11. ELLIOTT, M. H. The effect of appropriateness of rewards and of complex incentives on maze performance. *Univ. Calif. Publ. Psychol.*, 1929, **4**, 91–98.
12. ELLIOTT, M. H. Some determining

factors in maze performance. *Amer. J. Psychol.*, 1930, **42**, 315–317.

13. GLEITMAN, H. Studies in motivation and learning: II. Thirsty rats trained in maze with food but no water; then run hungry. *J. exp. Psychol.*, 1950, **40**, 169–174.
14. GRICE, G. R. An experimental test of the expectation theory of learning. *J. comp. physiol. Psychol.*, 1948, **41**, 137–143.
15. HANEY, G. W. The effect of familiarity on maze performance of albino rats. *Univ. Calif. Publ. Psychol.*, 1931, **4**, 319–333.
16. HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.
17. HAYES, K. J. An experimental criticism of the Spence and Lippitt procedure. *Amer. Psychologist*, 1949, **4**, 223. (Abstract.)
18. HEBB, D. O. The effects of early experience on problem-solving at maturity. *Amer. Psychologist*, 1947, **2**, 306–307. (Abstract.)
19. HERB, FRANCES H. Latent learning—non-reward followed by food in blinds. *J. comp. Psychol.*, 1940, **29**, 247–255.
20. HERON, W. T. Internal stimuli and learning. *J. comp. physiol. Psychol.*, 1949, **42**, 486–492.
21. HILGARD, E. R. *Theories of learning.* New York: Appleton-Century, 1948.
22. HULL, C. L. Goal attraction and directing ideas conceived as habit phenomena. *Psychol. Rev.*, 1931, **38**, 487–506.
23. HULL, C. L. The goal gradient hypothesis and maze learning. *Psychol. Rev.*, 1932, **39**, 25–43.
24. HULL, C. L. Differential habituation to internal stimuli in the albino rat. *J. comp. Psychol.*, 1933, **16**, 255–273.
25. HULL, C. L. The mechanism of the assembly of behavior segments in novel combinations suitable for problem solution. *Psychol. Rev.*, 1935, **42**, 219–245.
26. HULL, C. L. Conditioning: outline of a systematic theory of learning. *Yearb. nat. Soc. Stud. Educ.*, 1942, **41** (II), 61–95.
27. HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
28. KARN, H. K., & PORTER, J. M., JR. The effects of certain pretraining procedures upon maze performance and their significance for the concept of latent learning. *J. exp. Psychol.*, 1946, **36**, 461–469.
29. KENDLER, H. H. The influence of simultaneous hunger and thirst drives upon the learning of two opposed spatial responses of the white rat. *J. exp. Psychol.*, 1946, **36**, 212–220.
30. KENDLER, H. H. A comparison of learning under motivated and satiated conditions in the white rat. *J. exp. Psychol.*, 1947, **37**, 545–549.
31. KENDLER, H. H. An investigation of latent learning in a T-maze. *J. comp. physiol. Psychol.*, 1947, **40**, 265–270.
32. KENDLER, H. H. An experimental examination of the non-selective principle of association of drive-stimuli. *Amer. J. Psychol.*, 1949, **62**, 382–391.
33. KENDLER, H. H., & MENCHER, HELEN C. The ability of rats to learn the location of food when motivated by thirst—an experimental reply to Leeper. *J. exp. Psychol.*, 1948, **38**, 82–88.
34. KRECHEVSKY, I. A study of the continuity of the problem-solving process. *Psychol. Rev.*, 1938, **45**, 107–133.
35. LASHLEY, K. S. A simple maze: with data on the relation of the distribution of practice to rate of learning. *Psychobiology*, 1918, **1**, 353–367.
36. LASHLEY, K. S. Learning: I. Nervous mechanisms of learning. In C. Murchison (Ed.), *The foundations*

of experimental psychology. Worcester, Mass.: Clark Univ. Press, 1929. Pp. 524–563.

37. Lashley, K. S., & Wade, Marjorie. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.
38. Leeper, R. W. The role of motivation in learning: a study of the phenomenon of differential motivational control of the utilization of habits. *J. genet. Psychol.*, 1935, **46**, 3–40.
39. Leeper, R. W. Dr. Hull's *Principles of behavior*. *J. genet. Psychol.*, 1944, **65**, 3–52.
40. Leeper, R. W. The experiments by Spence and Lippitt on the sign-gestalt theory of learning. *J. exp. Psychol.*, 1948, **38**, 102–106.
41. Littman, R. A. Latent learning in a T-maze after two degrees of training. *J. comp. physiol. Psychol.*, 1950, **43**, 135–147.
42. Lorge, I. Irrelevant rewards in animal learning. *J. comp. Psychol.*, 1936, **21**, 105–128.
43. MacCorquodale, K., & Meehl, P. E. "Cognitive" learning in the absence of competition of incentive. *J. comp. physiol. Psychol.*, 1949, **42**, 383–390.
44. McGeoch, J. A. *The psychology of learning.* New York: Longmans, Green, 1942.
45. Maltzman, I. M. An interpretation of learning under an irrelevant need. *Psychol. Rev.*, 1950, **57**, 181–187.
46. Meehl, P. E. On the circularity of the law of effect. *Psychol. Bull.*, 1950, **47**, 52–75.
47. Meehl, P. E., & MacCorquodale, K. A further study of latent learning in the T-maze. *J. comp. physiol. Psychol.*, 1948, **41**, 372–396.
48. Meehl, P. E., & MacCorquodale, K. A failure to find the Blodgett effect and some secondary observations on drive conditions. *Amer. Psychologist*, 1949, **4**, 222. (Abstract.)
49. Miller, N. E. A reply to "Sign-gestalt or conditioned reflex?" *Psychol. Rev.*, 1944, **42**, 280–292.
50. Postman, L. The history and present status of the law of effect. *Psychol. Bull.*, 1947, **44**, 489–563.
51. Reynolds, B. A repetition of the Blodgett experiment on "latent learning." *J. exp. Psychol.*, 1945, **35**, 504–516.
52. Seward, J. P. A theoretical derivation of latent learning. *Psychol. Rev.*, 1947, **54**, 83–98.
53. Seward, J. P. An experimental analysis of latent learning. *J. exp. Psychol.*, 1949, **39**, 177–186.
54. Simmons, R. The relative effectiveness of certain incentives in animal learning. *Comp. Psychol. Monogr.*, 1924, **2**, 1–79.
55. Spence, K. W. Theoretical interpretations of learning. In F. A. Moss (Ed.), *Comparative psychology* (Rev. Ed.). New York: Prentice-Hall, 1942. Pp. 280–329.
56. Spence, K. W. An experimental test of the continuity and non-continuity theories of discrimination learning. *J. exp. Psychol.*, 1945, **35**, 253–266.
57. Spence, K. W. The role of secondary reinforcement in delayed reward learning. *Psychol. Rev.*, 1947, **54**, 1–8.
58. Spence, K. W., & Kendler, H. H. The speculations of Leeper with respect to the Iowa tests of the sign-gestalt theory of learning. *J. exp. Psychol.*, 1948, **38**, 106–109.
59. Spence, K. W., & Lippitt, R. "Latent" learning of a simple maze problem with relevant needs satiated. *Psychol. Bull.*, 1940, **37**, 429. (Abstract.)
60. Spence, K. W., & Lippitt, R. An experimental test of the sign-gestalt theory of trial and error learning. *J. exp. Psychol.*, 1946, **36**, 491–502.
61. Strange, J. R. Latent learning under conditions of high motivation. *J.*

comp. physiol. Psychol., 1950, **43**, 194–197.

62. Szymanski, J. S. Versuch über die Wirkung der Faktoren, die als Antrieb zum Erlernen einer Handlung dienen können. *Pflüg. arch. ges. Physiol.*, 1918, **171**, 374–385.
63. Thistlethwaite, D. An experimental test of a reinforcement interpretation of latent learning. *J. comp. physiol. Psychol.* (In press.)
64. Tolman, E. C. *Purposive behavior in animals and men.* New York: Appleton-Century, 1932.
65. Tolman, E. C. The law of effect. *Psychol Rev.*, 1938, **45**, 165–203.
66. Tolman, E. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948, **55**, 189–208.
67. Tolman, E. C. The nature and functioning of wants. *Psychol. Rev.*, 1949, **56**, 357–369.
68. Tolman, E. C. The psychology of social learning. *J. soc. Issues*, Suppl. Ser., 1949, No. 3.
69. Tolman, E. C. There is more than one kind of learning. *Psychol. Rev.*, 1949, **56**, 144–155.
70. Tolman, E. C., & Honzik, H. C. Introduction and removal of reward, and maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1930, **4**, 257–275.
71. Walker, E. L. Drive specificity and learning. *J. exp. Psychol.*, 1948, **38**, 39–49.
72. Walker, E. L., Knotter, Margaret C., & DeValois, R. L. Drive specificity and learning: the acquisition of a spatial response to food under conditions of water deprivation and food satiation. *J. exp. Psychol.*, 1950, **40**, 161–168.
73. Wallace, S. R., Jr., Blackwell, M. G., Jr., & Jenkins, I. Prereward and postreward performance in the "latent learning" of an elevated maze. *Psychol. Bull.*, 1941, **38**, 694. (Abstract.)
74. White, R. K. The case for the Tolman-Lewin interpretation of learning. *Psychol. Rev.*, 1943, **50**, 157–186.
75. Williams, Katherine A. The reward value of a conditioned stimulus. *Univ. Calif. Publ. Psychol.*, 1929, **4**, 31–55.
76. Young, P. T. Food-seeking drive, affective process and learning. *Psychol. Rev.*, 1949, **56**, 98–121.

Received May 19, 1950.

EXPERIMENTAL EVIDENCE FOR THE MOTIVATING FUNCTION OF REWARD

JOHN P. SEWARD
University of California, Los Angeles

INTRODUCTION

In a previous article (**33**) I proposed a theory to account for the action of reinforcers in a learning situation. According to that theory a reinforcer does not strengthen a habit as such but provides motivation for its use on future occasions. In this respect my position is like Tolman's; it may be considered an attempt to work out in more detail the mechanisms implied in such a view. Briefly, I suggest that an animal in a state of need is motivated not only by a primary drive (D) and drive stimulus (S_D) but by a secondary drive consisting of a set (rs_G) to make a characteristic consummatory or goal response (R_G). When a response (R) is followed by reward, R_G is conditioned to concurrent stimuli. By generalization of this conditioning, stimuli accompanying R now serve to intensify rs_G. It is further proposed that this intensification be called *tertiary motivation* and that it be endowed with the property of facilitating R, the activity in progress.

Our theory holds that a useful distinction may be drawn between two classes of learned responses, habits ($_sH_R$) and motives ($_sH_{rs_G}$). Both classes are covered by a general law of association of which Hull's Postulate 4 (**18**, p. 178) is a special case referring to $_sH_{rs_G}$. Reaction potential ($_sE_R$) is conceived of as a product of the following factors (in addition to stimulus intensity dynamism (**19**): D, $_sH_{rs_G}$, and $_sH_R$. This brief résumé is enough to indicate that, although the content of the theory stems from Tolman, its form is taken almost bodily from Hull.

The aim of the present article is to test the above formulation in the light of experimental evidence. To this end I have chosen, as particularly relevant, evidence as to the role of the incentive in learning situations. Specifically we shall consider experiments on the effects of the following factors: (1) properties of reward, such as amount and palatability; (2) changes in properties of reward; and (3) introduction of reward. One highly significant operation—removal of reward—must be omitted until our theory has been extended to include the facts of experimental extinction.

PROPERTIES OF REWARD

Simple Locomotor Response

Amount of reward. A number of studies (**5**, **10**, **39**, **43**) have shown that animals will approach a large reward more readily than a small one. Only two of these were sufficiently systematic and extensive to permit a test of Hull's quantitative statement (20), here adopted, that $_sE_R$ at any stage of learning is a positive exponential function of the square root of the weight of food reward ($\sqrt{w}$). Crespi (5) measured the mean running speed of different groups of rats down a 20-foot alley to varying amounts of food. Zeaman (43) also varied the amount of reward but measured the latent time of starting down a three-foot elevated pathway. In both studies increase of incentive was accompanied by improved performance. Moreover, if measures taken after improvement has ceased are plotted against $\sqrt{w}$, a curve suggestive of an exponential function is obtained.

Crespi was not primarily concerned with learning, though the three curves of improvement he presents seem by inspection to approach their asymptotes at the same rate. Zeaman, however, made a direct test of Hull's assumption that this rate of approach is independent of the incentive function. On fitting exponential equations to the mean log latencies during acquisition he showed that the constant controlling fractional increment per trial was practically constant. As Zeaman reminds us, these findings can be used to confirm Hull's hypothesis only by assuming that $_sE_R$ is proportional to performance measures; i.e., to running speed as measured by Crespi and inversely to log latency as measured by Zeaman. The latter assumption is not made by Hull. He makes $_sE_R$ a power function of latency (9, 18, 19), thereby implying that a linear relation to log latency could be approximated only over a limited range. But if Hull's mathematical postulates fail of exact confirmation, they are confirmed at the level of approximation which is more characteristic of his system in its present state (see 9, p. 525, footnote 8).

So far, Hull's revised postulate set appears adequate to account for the data. But Wolfe and Kaplon's (39) most striking discovery remains unexplained. They compared the performances of chicks as affected by one of three incentives: one large piece of popcorn (Group I), one quarter of a piece (Group II), or four quarters (Group III). Since the largest differences they found, on the whole, were between Group III and the other two groups, they concluded that amount of consummative activity was more effective than amount of reward.

The theory here proposed attributes the influence of an incentive to the arousal of a fractional antedating goal reaction, r_{S_G}. The more intensely r_{S_G} is aroused the greater will be its effect. At first glance, therefore, it seems reasonable to expect that four R_Gs in rapid succession will be more effective than one. Closer inspection reveals two ways in which such a result may be obtained: (1) It is possible that each R_G may function as if it were a separate trial and add an increment of habit strength to $_sH_{r_{S_G}}$; (2) the successive R_Gs on a single trial may summate to condition a more vigorous r_{S_G} to the situation. That is, repeating R_G may either condition rs_G more rapidly or condition a stronger rs_G. The two possibilities are open to experimental test. If Wolfe and Kaplon's finding were due to accelerated conditioning the performance curves of Groups I and III should approach the same limit but at different rates. If it was due to an intensified R_G the curves should approach different asymptotes, but their relative slopes would be the same. The data at hand do not permit a choice between these alternatives; the second one, however, seems more likely.

Either Hull's theory or the present revision can handle this interpretation quite readily. Hull's "incentive function" (20) need only be modified by substituting for $\sqrt{w}$ the construct R_G, thus:

$$K = 1 - 10^{-kR_G},$$

in which R_G stands for the intensity of the consummatory response. R_G in turn may be considered a function of several characteristics of the goal object; e.g., weight, size, nutritive value, palatability as measured in preference tests, and amount of consummative activity. The more precise form of this function must await the outcome of experiments in which these determinants are systematically varied.

Palatability of reward. Study of one of these dimensions, palatability, has already been carried some distance by Young (42). He defined palatability as "the relative acceptance of food [as] dependent upon the characteristics of the foodstuff" (42, p. 100). Within the general problem of how this property is related to behavior, Young asked himself the specific question: Does it influence learning or performance? In one experiment (40) he trained rats to run to sugar, wheat, or casein in counter-balanced order and found consistent differences in rate of running to the three incentives in successive cycles. His method of separating learning from performance was to equate the total times for a series of 30 runs and compare the numbers of runs in successive thirds of that total. The data for all three incentives, when treated in this way, turned out to be closely similar. Young concluded that although palatability was a factor in rate of running there was no evidence that it determined rate of learning.

Two features of Young's method limit the generality of his finding. (1) His rats had to shuttle back and forth between starting box and food; it is possible that if they had run one way only the results might have been different. In another connection Lawrence and Miller (24) showed how a similar change of technique might lead to an opposite conclusion. (2) His analysis was based on the first 30 runs to each incentive rather than on the total time required to reach an asymptote. His conclusion may therefore apply only to an early stage of practice. As far as they go, however, Young's data on palatability agree with those on amount in strengthening the view that rewards function mainly as motivators rather than "associators."

Spatial Discrimination

In one of Grindley's (10) experiments his chicks had to find their way to food through a simple maze and did so slightly faster to six grains of rice than to one. But no "error" (excess distance) scores were given. Similarly Wolfe and Kaplon (39) included a single-unit T-maze in their study; they reported a smaller error total for the group running to a multiple incentive than for either of the others. An experiment aimed directly at this problem was recently carried out by Reynolds (30), who compared two groups of rats in learning a single-unit T-maze to a 30-mg. pellet and a 160-mg. pellet, respectively. The large reward group proved consistently superior both in per cent correct choices and in mean running times. Young (41) blocked the preferred path in a Y-maze and compared both time and excess distance scores of rats running to sugar and to casein. His results are not strictly comparable with those of Reynolds, who used a non-correction method, but they point to the same general conclusion: that the increase in preference for a food path over a non-food path is a function both of amount and of palatability of reward.

Visual Discrimination

In another experiment, however, Reynolds (29) trained two groups of rats on a black-white discrimination problem, again using a 30-mg. and a 160-mg. pellet, respectively; but here he found no appreciable difference between them either in mean number of trials to a criterion or in per cent correct responses during learning.[1] Similar results for palatability were obtained by Young (41), who compared the rates of learning a black-white discrimination to sugar and to casein as rewards and found no difference in error scores. How can these negative findings be reconciled with the positive ones on locomotion and spatial discrimination?

Young's conclusion from his own results is much like the writer's. He held that the law of effect applies not to "ordinary" learning but to the acquisition of a food-seeking drive. As to *how* it applies he wrote: "Degree of effect is related not directly to the rate of acquisition but rather to the strength of drive which is acquired" (42, p. 300).

Reynolds made a highly similar interpretation but stated it more precisely; we shall therefore consider it here in some detail. It will be recalled that Hull (18), on the basis of Gantt's (8) and Grindley's (10) data, had made the limit of habit strength an exponential function of amount of reward. To account for his negative finding Reynolds proposed that this equation be "limited to responses which can be considered replications of the final consummatory response" (29, p. 768). According to his "R-hypothesis" Hull's equation simply means that different amounts of reward arouse consummatory responses of different degrees of strength to be conditioned. Obviously this suggestion, too, has much in common with mine. I, too, propose that Hull's Postulate 4 be restricted to the conditioning of the goal response or a fraction thereof and that his "incentive function" be taken to represent the strength of that response.[2]

---

[1] This experiment is open to some question on grounds of method. As the author points out, the low-reward group may possibly have been favored by an unequal elimination of slow learners and by use of a correction method, permitting differential reinforcement of wrong responses. We are here assuming that his results can be generalized beyond these conditions.

In its present form, however, there is room for doubt that the R-hypothesis leads to unequivocal predictions. The trouble seems to lie with the definition of terms. Reynolds defines R_G as "the complete pattern of activity commonly associated with the appearance of the reinforcing stimulus" (30, p. 152). He then goes on to define a learned replicate response as "one which can be measured on the same dimensions as the consummatory response." It is doubtful whether these definitions provide adequate criteria for distinguishing replicate from non-replicate responses. What about a bar-pressing response, for example? Learning a spatial discrimination presents similar difficulties. If running down a straightaway is a replication of R_G, then why not running down an alley with a right-angle turn in it? A T-maze thus presents a choice of two replicate responses; since one of them is reinforced it will be stronger than the other. But a large reward will increase this difference more than a small one, so the correct turn will be earlier and more consistently chosen. So far so good. But if, by the R-hypothesis, spatial discrimination is a function of amount of reward, the same prediction should apply to visual discrimination. Here the problem may be thought of as learning two replicate responses, one to white, the other to black; the rest of the argument is essentially the same.[3] Of course these interpretations do not *have* to be made; my point is that as long as they *can* be made the hypothesis runs the danger of "predicting" whatever results are obtained.

It should be clear that this criticism is directed at the definitions, not at their author's use of them. In practice Reynolds has confined the term "replicate response" to measures of locomotor speed. He himself derived his T-maze results not directly from the R-hypothesis but indirectly from the shorter delay of reinforcement produced by the larger reward. The question remains, however, why he did not make the same prediction for discriminating white from black, especially since his large-pellet group gave shorter response times throughout.

---

[2] Postulate 4 was confusing in that sH_R was actually a complex entity determined partly by strength of response and partly by strength of conditioning. In his revised postulate set (20) Hull has removed this source of confusion by defining sH_R strictly in terms of number of reinforcements and entering the determinants of *response* strength (amount and delay of reinforcement) separately in the equation for sE_R. Since my proposed revision retains the symbol H in its original capacity, the same purpose can be served by introducing a new symbol, such as sC_R, to denote strength of conditioning.

[3] Only a sketch is suggested here. A complete proof would have to take account of other factors; e.g., generalization and extinction in the non-correction method, delay of reward in the correction method.

To criticize is one thing; to suggest anything better is another. Before giving a precise meaning to R_G I should like to know what particular reaction, if any, is essential to "reinforcement," or tertiary motivation. Meanwhile I believe the term will be most useful if we confine it to actual commerce with the goal object; e.g., to eating. This usage will enable R_G and other habits, locomotor or otherwise, to vary separately. The effect of increased reward on any habit R will then depend on how it affects R in relation to other responses evoked by the situation. We are thus in a better position to reconcile the divergent findings on spatial and visual discrimination.[4]

The learning of a spatial discrimination has already been dealt with in terms of the present theory (**31**, **32**). Essentially it involves the building up of a stronger reaction potential to stimuli accompanying a right turn, say, than to those accompanying a left. That is, ${}_{S_R}E_{R_R} > {}_{S_L}E_{R_L}$, which in turn depends on the relation: ${}_{S_R}H_{rs_G} > {}_{S_L}H_{rs_G}$. If by increasing the intensity of rs_G the strength of ${}_{S}H_{rs_G}$ is doubled, the two competing ${}_{S}E_{R}$s and the absolute difference between them are also doubled. If we grant certain other assumptions this will result in a higher proportion of correct choices.

Applied to discriminations other than spatial, however, the above account is seen to be inadequate. Lawrence's recent evidence (**23**) points strongly to the existence of a "mediating process" in such discriminations, which apparently corresponds to some sort of perceptual organization. His results imply what field theorists have long maintained: that an essential problem for the animal is learning to which cues to respond. Reynolds' rats, for example, had not only to associate white with food but to *perceive* white in a sense that goes beyond merely seeing it. Or, to use a convenient notation, in order to acquire ${}_{S_W}H_{rs_G}$ it was also necessary to acquire ${}_{S}H_{rs_W}$.

In maze learning this mediating, perceptual process, though presumably present, has usually escaped attention. The reason is that rats generally respond to the important spatial cues, external and internal, from the beginning. But studies of place vs. response learning (**2**, **3**, **38**), by setting external and internal cues against each other, have brought this aspect into prominence. What is learned in a maze with food on the right, then, could be more accurately designated: $sH_{rs_R}H_{rs_G}$.

---

[4] The following discussion is aimed primarily at Reynolds' study. Young's method differed in several respects—low level of hunger, 24-hour spacing of trials, separation of cues from reward, scoring of time and errors—which may call for a different explanation.

Let us now return to B–W discrimination. If we assume, with Spence (34), that a rat starts with a position preference, say to the right, it is quite likely that after a few trials we shall still have the inequalities: $sH_{rs_R}H_{rs_G} > sH_{rs_L}H_{rs_G}$, and $s_RE_{R_R} > s_LE_{RL}$. But this is precisely the condition on which we have seen that learning a *spatial* discrimination depends. Furthermore, the greater the reward the greater will be the difference in sE_R, and the greater this difference the more quickly will the animal make his choice. One more assumption is that the less time an animal delays in the choice area, the less chance he will have to respond to S_W or S_B rather than to S_R or S_L. From this argument the conclusion follows that an increase of reward will *interfere* with the learning of sH_{rs_W}.

It may be objected that I have overshot the mark, since it looks as if Reynolds' low-reward group should have been superior, whereas actually the means of the two groups were about the same. It must be remembered, however, that the deterrent effect of a large reward applies only to the acquisition of sH_{rs_W}; its effect on the other half of the process, $_{rs_W}H_{rs_G}$, will be just the opposite. Moreover, some rats in both groups may have weak position preferences and respond to S_W early in practice; of these the ones in the high-reward group should learn most rapidly of all. On the empirical side two characteristics of Reynolds' data may be cited in support of this interpretation: (1) the shorter mean response times for the high-reward group; (2) the greater variation and skewness of its learning score distribution (on the assumption that the tail of the distribution is contributed by a few rats with extremely persistent position habits).[5]

CHANGES IN PROPERTIES OF REWARD

Amount of reward. Besides comparing animals given different amounts of reward, several of the above investigators studied the effects of varying the amount of reward given the same animal. Wolfe and Kaplon, for example, at the close of their runway experiment gave the group trained to one small piece of popcorn two extra sessions with four pieces. "Their time scores immediately approximated those of group three" (previously trained to four pieces) (39, p. 358). Crespi and Zeaman measured the effect of suddenly increasing or decreasing amount of reward after a group had reached its asymptote and reported not only an abrupt change in level of performance but a "contrast effect." Crespi found that rats shifted to 16 units either upward or downward, overshot the mark as compared with a group run to 16 units from the beginning. Zeaman found the same effect, but in his experiment it was significant only for a relatively large shift and then only in the upward direction.[6]

---

[5] Actually the significance of these statistics for the point at issue is based on the assumption that there is a positive interaction between strength of initial position preference and the greater reinforcing effect of a large as compared with a small pellet.

Two features of the contrast effect require explanation: (1) the abruptness with which the new level was reached, and (2) its discrepancy from that of the control group (**5**, Figs. 7 and 8; **43**, Figs. 2 and 5). As previously pointed out (**33**), Hull's 1943 postulate set provided for neither of these phenomena. His 1949 revision (**20**), we noted, allows for rapid changes in $_sE_R$ with changes in K, the incentive function, but does not account for the learning of K; neither does it predict contrast. Crespi's own theory, that incentive-value, as one factor in motivation, is proportional to the distance between levels of expectation and attainment, is a stimulating suggestion; its further implications remain to be developed.

Sudden shifts. The present theory, it will be recalled, attempts to remedy the defect in Hull's treatment by using his Postulate 4 to describe the acquiring of a goal-expectation, $_sH_{rs_G}$. But if $_sH_{rs_G}$ has to be acquired by Postulate 4, how can we explain the one-trial shift of performance level when reward is increased or diminished? This result depends on the fact that rs_G has already been conditioned on the afferent side to the starting box at an intensity corresponding to the original amount of food, S_G. Denoting the starting box by S_S we have: $_{s_S}H_{rs_G}>0$. On the critical trial the changed incentive, $S_{G'}$, arouses $rs_{G'}$ of proportional intensity; i.e., $_{s_{G'}}H_{rs_{G'}}>0$. Since S_G and $S_{G'}$ are on a continuum of stimulus magnitude their central counterparts, rs_G and $rs_{G'}$, may be said to lie on a continuum of afferent intensity. By the principle of stimulus generalization, therefore, $_{s_{G'}}H_{rs_{G'}}$ should give a degree of effective habit strength to the connection $_{rs_G}\overline{H}_{rs_{G'}}$. Moreover, there is experimental evidence (**14**) that the gradient of primary stimulus-intensity generalization is quite flat. We may infer that on the next trial after the change of reward a process, $_{s_S}H_{rs_G}\overline{H}_{rs_{G'}}$, will be activated in the starting box and that it will be of comparable intensity with $_{s_S}H_{rs_{G'}}$ aroused in a group trained to $S_{G'}$ from the beginning. From this state of affairs the sudden change in latency and speed follows.

Contrast effect. An adequate theory of contrast effect is much too complex to be attempted here, but a few suggestions may be given. Assumptions are required whereby a change in magnitude of incentive will produce an increment or decrement of motivation beyond that due to a constant amount. We may begin by assuming that the consumption of any quantity of food short of satiation involves some frustration of the unsatisfied need (F_{R_G}); the less food, the more frustration. As the animal's set for this quantity of food is generalized to the starting box of a runway, a corresponding degree of frustration (F_{rs_G}) will also be aroused there and will act as a negative drive to oppose the positive drive to food.[7] For any given amount of incentive $S_{G'}$ we may compute the

---

[6] A second anomaly reported by Crespi—the fact that rats reduced from 16 units to no reward ran faster than rats reduced to 1 unit—may be an artifact, since the former group had 14 more runs to 16 units before the incentive was removed.

effective strength of tertiary motivation by substituting values in the following equation:

$$s\overline{H}_{rsG'} = sH_{rsG'} - F_{rsG'}. \qquad [1]$$

Assuming that the downward contrast effect is real, we may deduce it by supposing that F_{R_G} is also produced when a portion of R_G is blocked by a sudden reduction in the quantity of food. The greater the reduction, the stronger will be the frustration from this source. It is this added frustration that pulls down the effective motivation of the lowered group *below* that of the group already running to the smaller incentive.

To illustrate, let us suppose that a large quantity of food, $S_{G''}$, is suddenly reduced to $S_{G'}$. Through stimulus generalization, by the mechanism set forth above in deducing sudden shifts, rs_G will be aroused in the starting box with an intensity roughly equivalent to $rs_{G'}$, accompanied by $F_{rsG'}$. In addition there will be generated a quantity of F which is some function of the difference between $S_{G''}$ and $S_{G'}$ and which we may designate $F_{rs(G''-G')}$. In this case effective tertiary motivation is determined as follows:

$$s_S\overline{H}_{rsG'} = s_SH_{rsG''}\overline{H}_{rsG'} - [F_{rsG'} + F_{rs(G''-G')}]. \qquad [2]$$

To explain upward contrast, an analogous assumption must be made. If a frustrated animal is suddenly given access to food, some of his F_{R_G} will be channeled, or converted, into R_G. Similarly, when S_G is increased the same process will occur. This means that at the start of the next trial a corresponding part of F_{rsG} will be converted into sH_{rsG}; the greater the proportion of drive converted, the greater the contrast effect. In terms of equation [2] this operation on F_{rsG} is mathematically equivalent to a change of sign. Thus, if S_{G° represents a small amount of food suddenly increased to $S_{G'}$, we may write:

$$s_S\overline{H}_{rsG'} = s_SH_{rsG^\circ}\overline{H}_{rsG'} - [F_{rsG'} - F_{rs(G'-G_o)}]. \qquad [3]$$

The foregoing assumptions are of little value, of course, until they can be put to work; i.e., until the concepts involved are quantitatively related to observables. Such statements in turn depend on more studies like Zeaman's, in which both degree of incentive change and drive level are systematically varied and performance measures plotted against them.

Palatability of Reward. Needless to say, our interpretation of contrast is not intended to be purely ad hoc. To defend it from that charge we should at least be prepared to show that it applies to other dimensions besides the one that gave rise to it. Many years ago Elliott (7) disrupted the maze performance of rats by changing their reward from bran mash to sunflower seed. The effect was sudden and, in comparison with a control group trained to sunflower seed from the beginning, showed a marked "contrast effect" which the author attributed to searching for mash. More recently Young (41) tested the effect of interchanging rewards of different degrees of palatability. Unfortunately his study serves mainly to underline the hazards of gen-

---

[7] We could arrive at the same result by assuming that F_{R_G} is conditioned to the starting box along with food anticipation, thus becoming sF_{rsG}. Although not indispensable here, this assumption may prove to be a decided advantage in dealing with extinction.

eralizing in this area. Young gave his rats one trial a day to a nibble of food on an open platform and measured their approach time. One group ran to an incentive of high palatability (sugar), another to a lower one (casein). He then shifted half of his animals to the opposite incentive. Under these conditions there was no contrast effect but a difference appeared between the two directions of shift. The upward shift produced an immediate reduction of times, but the downward shift produced no retardation for five days. Whether we call the second phenomenon "wishful thinking" or simply lag, it looks like just the opposite of contrast.

It would be foolhardy to attempt an explanation here. (One ad hoc hypothesis at a time!) But before attributing the discrepancies to a basic difference between amount and palatability of reward we should make sure that they are not due to a difference in the conditions of measurement; e.g., time of food deprivation, type and length of runway, method of scoring, or distance between scaled incentive values. Nor should the possibility of an interaction between amount and palatability be overlooked. Further investigations may show that certain conditions favor contrast and others favor lag, and that these relations are common to all properties of reward.

INTRODUCTION OF REWARD

Evidence on this topic comes from the well-known experiments on latent learning. In a recent excellent review of these experiments Thistlethwaite (36) classified them in four types, of which we are here concerned with the first two. In type I, the Blodgett variety, a group of hungry rats is given a series of trials in a maze with no food in the end-box, followed by a series with food. Evidence for latent learning consists of a sudden drop in errors with the introduction of food in comparison with a control group rewarded with food from the outset. In type II, the Haney variety, the rats are allowed to explore an empty maze for a time before being given regular trials to food. Latent learning is said to have occurred if their initial error and time scores are better than those of a control group allowed to explore a continuous runway.

Positive evidence from both types of experiment has led cognition theorists to the following conclusions, recently upheld by Thistlethwaite after a critical review of the entire field: (1) Habits may be acquired without being "reinforced" by "satisfiers." (2) Habits may be acquired without being reflected in overt performance. To these may be added a third: (3) "Satisfiers" determine the overt performance of habits rather than their acquisition.

All three of these propositions are implicit in the writer's tertiary-motivation theory of reinforcement. In fact, the theory was devised largely to meet the demands of the latent learning evidence. To evaluate its success in doing so the reader may consult earlier statements (**31**, **32**, **33**). Here we shall be concerned with whether the attempt was justified or whether an orthodox reinforcement theory would do as well.

Advocates of what Meehl (**26**) calls the "Strong Law of Effect" answer the challenge of latent learning in a number of ways:

In answer to proposition (1) above, they point out that the pre-reward period is not devoid of satisfiers. In type I experiments the end-box, as the place of removal from the maze, may acquire secondary reinforcing properties by association with a subsequent anxiety-reducing environment. In both types the maze as a whole may arouse—and satisfy—a rat's need to explore. Whatever the mechanism of exploratory behavior, it seems to be evoked in situations having a common feature, novelty, and to persist until that feature has been eliminated. We cannot therefore deny it status as a need-reducing activity; all we can say in reply is that exploratory satisfactions should be distributed all over the maze instead of being concentrated in certain sections. But it is also possible that a rat may prefer a "through road" to a blind alley, so that even during free wandering he will enter fewer and fewer culs.

In answer to proposition (2) they assert that much so-called latent learning is actually "manifest" (**1**). The possibilities mentioned above are not sheer speculation. All the type I experiments reported have found some reduction of errors before the introduction of food. Karn and Porter (**21**) and Reynolds (**28**) showed how refined observations might detect evidence of learning normally overlooked. Though type II experiments usually provide no measure of such a trend, Haney's (**11**) data strongly suggest that it existed, since *on the first trial*, before they had ever encountered food in the maze, his experimental rats made less than half as many errors as the controls. One of Buxton's (**4**) control groups, tested without pre-feeding, confirmed this finding. Finally, MacCorquodale and Meehl (**25**), using Blodgett's six-unit maze in a direct attack on the question, demonstrated a significant reduction in cul entries from a first to a second exploratory period and to an initial test run.

We must conclude that both "reinforcement" and "improved" performance characterize the pre-reward period. Introduction of reward is actually a change in the quantity and quality of need satisfactions. Yet the reinforcement theorist still has to explain the sharp elimination of errors which sometimes accompanies that change. It is another form of the same problem we faced in connection with changing the amount or palatability of reward.

A third way of trying to solve this problem in Hullian terms is to add more postulates to the system. Thistlethwaite (36) has already disposed of one such suggestion by Meehl by showing that it could not account for Herb's (13) data. I should like to consider three more attempts of this sort that have come to my attention:

1. Thistlethwaite's criticism also applies to the hypothesis recently offered by Meehl and MacCorquodale (27). Their argument starts with the fact of pre-reward improvement, which they interpret as indicating a weak difference in sH_R in favor of the true path. They assume further that eating food in the maze conditions drive-mediating responses to maze cues. As a result this conditioned drive, multiplied by accumulated sH_R, increases the differences in sE_R already existing at choice points. The authors tested their hypothesis by substituting for the first rewarded trial in Blodgett's maze A, two minutes of feeding in an extra-maze box. On the next trial the drop in errors, as compared with a control group put in the box without feeding, fell just short of a .05 level of significance.[8]

From my standpoint this interpretation has much to recommend it; indeed, if its authors specified more clearly the nature of the secondary motivation they invoked, their theory might prove to be essentially like mine. In its present form, however, it relies simply on the multiplication of existing habit differentials and thus fails to do justice to Herb's results. That investigator, it will be remembered, modeled her experiment on that of Tolman and Honzik (37), but when she introduced food in the maze she put it at the ends of the blind alleys instead of in the endbox. During the pre-reward trials, while the control group entered more and more blinds the experimental group entered fewer and fewer. Nevertheless, following the introduction of food the experimental group increased its blind-alley entrances significantly more than the control group in any equivalent period.

This abrupt change of slope *in the opposite direction* strongly implies that the previous learning was something other than a differential strengthening of overt responses at successive choice points. It suggests rather that what was learned at any choice point was an antedating reaction to a "dead end" (rs_E) in one direction and to a "through road" (rs_C) in the other. We may assume further that the chief source of differential motivation was rs_G and that in the absence of food this consisted of a return to the home cage or of unimpeded activity. The early reduction of errors may then be accounted for by either or both of two processes: the conditioning of tertiary motivation to rs_C and of frustration to rs_E (e.g., ${}_{S_R}H_{rs_C}H_{rs_G}$ and ${}_{S_L}H_{rs_E}F_{rs_G}$). But the discovery of food in a blind alley should condition a more powerful rs_G to the rs_E aroused at that choice point, thus tipping the balance in favor of entering the blind. By virtue of stimulus generalization the same process should occur to a lesser degree at other choice points and lead to further discoveries of food. Thus we should expect the rapid increase of "errors" which, in fact, took place.

---

[8] Personal communication.

2. Another attempt to buttress Hull's theory with qualifying assumptions was made by Taylor (35). He took as his point of departure my analysis of Blodgett's rectangular-maze data (31). My treatment had a number of faults; for one thing, I failed to take account of the fact that Blodgett's experimental group developed a slight preference for the short path during the pre-food period. Taylor's main objection, however, was on the ground of parsimony, since Hull's system, with the aid of one or more propositions, could be shown to predict Blodgett's results more efficiently than mine. His achievement hinged on the assumption that behavioral oscillation (σ_0) decreases as $_sH_R$ increases. But in order to reconcile this assumption with other data Taylor took the position that it holds only under certain conditions. Specifically he suggested that when $_sH_R$ undergoes primary reinforcement, σ_0 decreases, but when reinforcement is secondary, σ_0 remains constant. This assumption seems to me untenable for the following reasons:

In the first place Taylor attributed primary reinforcement to the diminution of a primary or unlearned need such as hunger. Yet as examples of decreased variability with practice he used the skilled typist, the expert pianist, and a small group of subjects who were paid for their services in a study of autokinetic movement.

Secondly, Taylor based his argument on the ground that oscillation is one expression of drive and that, therefore, a reduction of primary drive will lead to a reduction of oscillation. This effect becomes conditioned to the drive-reducing situation. "Thus when a hungry animal finds itself in a situation where it has previously discovered food, there is likely to be an immediate reduction in general irritability" (35, p. 84). But this description of the conditions determining a decreased σ_0 could serve equally well as an example of *secondary* reinforcement as that term is commonly defined (18). Primary and secondary reinforcement, as already noted elsewhere (33), are so intimately related that they have so far defied any clear-cut separation. But without an objective criterion for distinguishing between them Taylor's assumption becomes purely ad hoc.

Thirdly, although Taylor himself nowhere defines secondary reinforcement explicitly, he implies that he means by it the diminution of a secondary or learned drive (35, p. 85). But if range of oscillation is a function of strength of drive it should vary with a secondary drive as well as with a primary one. In effect, a drive is a drive whether aroused reflexly or by association. In deriving Blodgett's results Taylor assigned a constant σ_0 to the experimental group before introduction of reward and a decreasing σ_0 thereafter. Since the experimental group in the early trials was admittedly diminishing its exploratory drive it is hard to see any logical ground for the distinction.

The question remains whether my own theory can be amended to explain the pre-reward as well as post-reward improvement in Blodgett's experimental group. As already suggested, this can be done most simply by recognizing a weak incentive (rs_{G°)—e.g., escape from maze—associated with the endbox. Let us denote the short path by X and the long path by Y. As a result of differential delay, ${}_{S_X}H_{rs_{G^\circ}}$ will be greater than ${}_{S_Y}H_{rs_{G^\circ}}$, thus producing a greater ${}_{S}E_{R}$ in favor of the shorter path. But since rs_{G° is weak the difference will be slight, and the overlapping of oscillation distributions will keep the percentage of short-path choices ($\%X$) not far above 50. When food is introduced and a much stronger incentive ($rs_{G'}$) is thus associated directly with the endbox and indirectly with the choice point, the difference ${}_{S_X}E_{R_X} - {}_{S_Y}E_{R_Y}$ is abruptly increased. On the assumption that σ_0 remains unaltered, $\%X$ will also jump.

Of course that assumption may not be valid. My criticisms of Taylor's position were directed at certain inconsistencies; they leave his main suppositions untouched. The conditions determining σ_0 are almost wholly unexplored (18, p. 314). Quite aside from any distinction between primary and secondary motivation, Taylor's idea that σ_0 is a positive function of drive may prove to be correct. In that case the final step of our argument would not follow, since if σ_0 increased, for example, proportionally with ${}_{S}E_{R}$, the X and Y distributions would still overlap to the same degree.

It would still be possible, however, to escape from this dilemma. A way out is suggested by Spence's (34) theory of discrimination learning. Spence held that the excitatory strength of the stimulus complex on either side of a discrimination apparatus is the sum of the excitatory tendencies of the component stimuli. The discrimination is learned when the excitatory difference between the critical cues becomes great enough to outweigh all other differences that might arise in the situation.

In much the same fashion we may assume that for Blodgett's experimental group during the early trials the endbox (S_{G°) was only one of an indefinite number of competing incentives, relatively equal in strength, some favoring a right turn and others a left. Due to oscillation, ${}_{S}H_{rs_{G^\circ}}$ won out in the competition only often enough to produce a small majority of short-path choices. But after food ($S_{G'}$) had been discovered it furnished such a strong incentive that its oscillation hardly overlapped with any of the others; ${}_{S}H_{rs_{G'}}$ therefore almost invariably dominated the animal's choice. Essentially this amendment would suggest that $\%X$ without food is largely a function of the relative potencies of competing incentives; with food in the endbox $\%X$ is largely a function of the difference between the ${}_{S}E_{R}$s associated with the two means of getting there.

3. Finally I should like to discuss Hull's own attempt (17) to tailor his theory to fit latent learning. Buxton (4) had recently published

his disconcerting monograph. Using a type II design combined with a brief pre-feeding in the endbox, he had demonstrated for his experimental groups not only less errors than for the controls but a speed-of-locomotion gradient and a backward order of error elimination *on the very first trial*. Since these phenomena had been deduced from—and accepted as evidence for—Hull's goal gradient hypothesis (15), here was the most formidable challenge to reinforcement principles that had yet appeared.

In 1941 Hull issued a memorandum (17) containing a proof of two propositions embodying two of Buxton's results: the smaller number of errors on the first trial by his experimental rats and the diminishing error gradient as they approached the goal. To the six postulates in his 1937 miniature system he added three assumptions about maze behavior: (1) that animals make perceptual reactions to their environment; (2) that exploratory behavior leads to the reinforcement of habit segments; and (3) that these segments are distributed about the maze by chance. Hull's proof is too intricate to be reproduced here. It follows essentially the same line of argument that he developed (16) in deducing Maier's results on "reasoning," of which he considered Buxton's problem to be a special case. To condense the exposition still further, however, it will be convenient to use my simplified version (32) of the Buxton experiment. That study may be broken down into three stages. In stage 1 the experimental rats were allowed to explore an enclosed single-unit T-maze with choice point *C* and endboxes *A* and *B* to right and left, respectively. In stage 2 they were put directly into one of the two endboxes, say *A*, where they found food for the first time. In stage 3, which followed immediately, they were placed in the starting box and given a free choice; 87.5 per cent chose the path to *A*.

Hull's deduction of this finding, boiled down from his discussions of Maier and Buxton, would run somewhat as follows (in the accompanying diagrams a broken arrow indicates an association in the making; a solid arrow, an association already formed):

1. As a result of maze exploration fractional reactions to the endboxes, r_A and r_B, will be brought forward to the choice point and their proprioceptive impulses, s_A and s_B, conditioned to the respective approach responses, R_R and R_L.

$$\begin{tikzcd}
 & r_B \arrow[r] & s_B \arrow[r, dashed] & R_L \\
\text{STAGE 1:}\quad S_C \arrow[ur, dashed] \arrow[dr, dashed] & & & \\
 & r_A \arrow[r] & s_A \arrow[r, dashed] & R_R
\end{tikzcd}$$

2. Pre-feeding ($S_G - R_G$) at *A* conditions the hunger drive stimulus, S_D, to endbox responses, R_A.

STAGE 2:

$$\begin{tikzcd} & S_A \arrow[r] & R_A \\ S_D \arrow[ur] \arrow[dr] & & \\ & S_G \arrow[r] & R_G \end{tikzcd}$$

3. When the animal reaches the choice point a few seconds later, r_A is still perseverating and is now strongly aroused by S_D. R_R is therefore more intensely activated than R_L.

STAGE 3:

$$\begin{tikzcd} & r_B \arrow[r] & s_B \arrow[r] & R_L \\ S_C \arrow[ur] \arrow[dr] & & & \\ & {\left.\begin{array}{r} r_A\rightarrow s_A \\ S_D\rightarrow r_A\rightarrow s_A \\ r_A\rightarrow r_A\rightarrow s_A \end{array}\right\}} \arrow[r] & R_R & \end{tikzcd}$$

By applying the same reasoning to a succession of choice points in conjunction with his third supplementary assumption, Hull was able to deduce the main features of Buxton's error data. Although his system has advanced far beyond the stage it was in at that time, I assume that his main argument still holds. It rests on a number of characteristic assumptions, of which we may note the following: (a) Learning is a function of satisfiers. (b) Other things being equal, selection between incompatible responses is a function of relative habit strength. (c) Total ${}_sH_R$ is a function of ${}_sH_R$ components. (d) The role of S_D in selection is to add a stronger ${}_sH_R$ component to one response than another.

My own interpretation resembles Hull's up to a point. It includes the conditioning of endbox reactions to choice-point stimuli in stage (1). In stage (2), however, it stresses the association of endbox reactions with eating responses (R_G) rather than with S_D. And in stage (3) it states that the differential arousal of rs_G thus produced at the choice point increases ${}_sE_R$ in favor of the appropriate response. Learning, according to this view, may be largely independent of satisfiers. Selection is determined not so much by relative ${}_sH_R$ as by relative intensity of tertiary motivation, ${}_sH_{rs_G}$. The chief role of S_D is to establish a set toward R_G, which is then differentially blocked or facilitated by external and response-produced—and surrogate-response-produced—stimulation.

But are the two theories actually different in the sense that they lead to different predictions? Since Hull's proof hinges on the greater strength of S_A than S_B, it follows that any means of making it stronger should produce a like result. For example, the starting box could be made more similar to box A. Again, it should be enough merely to put the rat in A without food just before making a choice; the perseverative trace of s_A should then favor a right turn. Evidence against this hypothesis was furnished by another part of my experiment. When rats were put alternately in both boxes just before testing, half of them fed in the first box and half in the second, 81.3 per cent still chose the food side. Apparently sheer perseveration contributed little if anything to the result. Further negative evidence is the well-known disposition of rats to alternate their responses at a choice point, a tendency that is stronger the shorter the time interval between trials (6, 12).

The second factor in Hull's proof contributing to s_A is the conditioning of S_D to r_A. As long as this is accomplished by reducing S_D itself, R_G is necessarily involved and both theories yield the same prediction. But what if some other need is reduced in the presence of S_D? What if the rats, still hungry, are put in A and permitted to slake their thirst? According to Hull's interpretation, 24 hours later the rats, now hungry but not thirsty, will nevertheless show a preference for A; according to mine they will not.[9] To avoid this deduction Hull would have to supplement his theory with Kendler's "selective principle" (22), according to which a drive stimulus is conditioned only by virtue of its own reduction. A discussion of the present status of this principle would carry us too far afield.

To summarize, we see that Hull has tried to explain latent learning on the assumption that satisfiers determine behavior solely by producing differences in $_sH_R$. In my opinion a more satisfactory solution can be found by assuming that a response to a satisfier, rs_G, is itself learned much like any other response; once learned it enters the stimulus complex as one form of motivation to help decide which of two $_sE_R$s will find overt expression.

SUMMARY

The present paper has been written to test the writer's tertiary-motivation hypothesis of reinforcement by applying it to an area of research. As the distinguishing feature of this area the independent variable was some property of reward, such as the presence or absence, amount, or palatability of a food incentive. Specifically, results of experiments on the following problems were discussed with respect to their theoretical significance:

1. *Effect of amount and of palatability of reward.* Concerning simple locomotor response, on the basis of Crespi's, Zeaman's, and Young's data, it was concluded that the incentive determines performance level at the limit of improvement but not the rate of approach to that limit. In the case of spat[illegible] discrimination, both time and error scores in mastering a choice point ha[illegible] yielded consistently positive results. For visual discrimination results here

---

[9] The original experiment would, of course, have to be repeated to test the effect of a 24-hour interval between stages 2 and 3.

have been negative. Reynolds' R-hypothesis was criticized, and another attempt was made to reconcile the findings on spatial and visual discrimination with Lawrence's evidence for a "mediating process" taken into consideration.

2. *Effect of changing the amount or palatability of reward on simple locomotor response.* Three behavioral phenomena were noted—sudden shifts, contrast effect, and lag—and interpretations of the first two were suggested. Sudden shifts were explained by means of an application of the principle of stimulus generalization to reward dimensions. To account for contrast effect an intervening variable of frustration was introduced; its quantity was assumed to vary as a function of both amount of reward and degree of change in that respect.

3. *Effect of introducing reward on maze performance.* Attempts to interpret latent learning data in terms of reinforcement theory were evaluated, especially those of Meehl and MacCorquodale, Taylor, and Hull. Implications of the writer's hypothesis were suggested as an alternative in each case.

BIBLIOGRAPHY

1. Blodgett, H. C. Reynolds' repetition of Blodgett's experiment on latent learning. *J. exp. Psychol.*, 1946, **36**, 184–186.
2. Blodgett, H. C., & McCutchan, K. Place versus response learning in the simple *T*-maze. *J. exp. Psychol.*, 1947, **37**, 412–422.
3. Blodgett, H. C., & McCutchan, K. Relative strength of place and response learning in the T maze. *J. comp. physiol. Psychol.*, 1948, **41**, 17–24.
4. Buxton, C. E. Latent learning and the goal gradient hypothesis. *Contr. psychol. Theory*, 1940, **2**, No. 6.
5. Crespi, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.
6. Dennis, W. Spontaneous alternation in rats as an indicator of the persistence of stimulus effects. *J. comp. Psychol.*, 1939, **28**, 305–312.
7. Elliott, M. H. The effect of change of reward on the maze performance of rats. *Univ. Calif. Publ. Psychol.*, 1928, **4**, 19–30.

[illegible]tt, W. H. The nervous secretion of saliva: the relation of the conditioned reflex to the intensity of the unconditioned stimulus. *Amer. J. Physiol.*, 1938, **123**, 74–75.

9. Gladstone, A. I., Yamaguchi, H. G., Hull, C. L., & Felsinger, J. M. Some functional relationships of reaction potential ($_sE_R$) and related phenomena. *J. exp. Psychol.*, 1947, **37**, 510–526.
10. Grindley, G. C. Experiments on the influence of the amount of reward on learning in young chickens. *Brit. J. Psychol.*, 1929, **20**, 173–180.
11. Haney, G. W. The effect of familiarity on maze performance of albino rats. *Univ. Calif. Publ. Psychol.*, 1931, **4**, 319–333.
12. Heathers, G. L. The avoidance of repetition of a maze reaction in the rat as a function of the time interval between trials. *J. Psychol.*, 1940, **10**, 359–380.
13. Herb, F. H. Latent learning-non-reward followed by food in blinds. *J. comp. Psychol.*, 1940, **29**, 247–256.
14. Hovland, C. I. The generalization of conditioned responses. II. The sensory generalization of conditioned responses with varying intensities of tone. *J. genet. Psychol.*, 1937, **51**, 279–291.
15. Hull, C. L. The goal gradient hypothesis and maze learning. *Psychol. Rev.*, 1932, **39**, 25–43.
16. Hull, C. L. The mechanism of the

assembly of behavior segments in novel combinations suitable for problem solution. *Psychol. Rev.*, 1935, **42**, 219–245.

17. Hull, C. L. Memorandum to Psychology 106. In *Psychological memoranda*, 1940–1944. Bound mimeographed material on file in the libraries of the University of Iowa, University of North Carolina, and Yale University. Nov. 14, 1941.
18. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
19. Hull, C. L. Stimulus intensity dynamism (V) and stimulus generalization. *Psychol. Rev.*, 1949, **56**, 67–76.
20. Hull, C. L. Behavior postulates and corollaries—1949. *Psychol. Rev.*, 1950, **57**, 173–180.
21. Karn, H. W., & Porter, J. M., Jr. The effects of certain pre-training procedures upon maze performance and their significance for the concept of latent learning. *J. exp. Psychol.*, 1946, **36**, 461–469.
22. Kendler, H. H. An experimental examination of the non-selective principle of association of drive-stimuli. *Amer. J. Psychol.*, 1949, **62**, 382–391.
23. Lawrence, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimuli. *J. exp. Psychol.*, 1949, **39**, 770–784.
24. Lawrence, D. H., & Miller, N. E. A positive relationship between reinforcement and resistance to extinction produced by removing a source of confusion from a technique that had produced opposite results. *J. exp. Psychol.*, 1947, **37**, 494–509.
25. MacCorquodale, K., & Meehl, P. E. On the elimination of cul-entries without obvious reinforcement. *J. comp. physiol. Psychol.* (In press.)
26. Meehl, P. E. On the circularity of the law of effect. *Psychol. Bull.*, 1950, **47**, 52–75.
27. Meehl, P. E., & MacCorquodale, K. A failure to find the Blodgett effect and some secondary observations on drive conditioning. *Amer. Psychologist*, 1949, **4**, 222. (Abstract.)
28. Reynolds, B. A repetition of the Blodgett experiment on latent learning. *J. exp. Psychol.*, 1945, **35**, 504–516.
29. Reynolds, B. The acquisition of a black-white discrimination habit under two levels of reinforcement. *J. exp. Psychol.*, 1949, **39**, 760–769.
30. Reynolds, B. Acquisition of a simple spatial discrimination as a function of the amount of reinforcement. *J. exp. Psychol.*, 1950, **40**, 152–160.
31. Seward, J. P. A theoretical derivation of latent learning. *Psychol. Rev.*, 1947, **54**, 83–98.
32. Seward, J. P. An experimental analysis of latent learning. *J. exp. Psychol.*, 1949, **39**, 177–186.
33. Seward, J. P. Secondary reinforcement as tertiary motivation: a revision of Hull's revision. *Psychol. Rev.*, 1950, **57**, 362–374.
34. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
35. Taylor, J. C. Behavior oscillation and the growth of preference. *Psychol. Rev.*, 1949, **56**, 77–87.
36. Thistlethwaite, D. L. A critical review of latent learning and related experiments. *Psychol. Bull.*, 1951, **48**, 97–129.
37. Tolman, E. C., & Honzik, C. H. Introduction and removal of reward, and maze performance in rats. *Univ. Calif. Publ. Psychol.*, 1930, **4**, 257–275.
38. Tolman, E. C., Ritchie, B. F., & Kalish, D. Studies in spatial

learning. II. Place learning versus response learning. *J. exp. Psychol.*, 1946, **36**, 221–229.

39. Wolfe, J. B., & Kaplon, M. D. Effect of amount of reward and consummative activity on learning in chickens. *J. comp. Psychol.*, 1941, **31**, 353–361.

40. Young, P. T. Studies of food preference, appetite and dietary habit. VII. Palatability in relation to learning and performance. *J. comp. physiol. Psychol.*, 1947, **40**, 37–72.

41. Young, P. T. Studies of food preference, appetite and dietary habit. VIII. Food-seeking drives, palatability and the law of effect. *J. comp. physiol. Psychol.*, 1948, **41**, 269–300.

42. Young, P. T. Food-seeking drive, affective process, and learning. *Psychol. Rev.*, 1949, **56**, 98–121.

43. Zeaman, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

Received June 1, 1950.

COMMENTS ON RAZRAN'S DISCUSSION OF STIMULUS GENERALIZATION

G. ROBERT GRICE
University of Illinois

Razran recently has made a valuable contribution by his summary of the literature on stimulus generalization (5). However, there are several points in Razran's interpretative discussion with which the writer is in disagreement and which he feels are in need of further clarification.

1. In the first place, Razran appears to be unduly concerned with the effect of variability upon empirically obtained generalization gradients. For example, with respect to Hovland's study, he says, "*Thus, not a single* [*one*] *of Hovland's 20 subjects revealed a consistent gradient of CR generalization and only one of the three gradient-determining differences was statistically reliable.*" The writer does not mean to minimize the importance of variability in behavior theory. However, the fact that variability may be superimposed upon underlying lawful relationships is in no way an argument against the existence of such relationships. This fact of behavioral variability has been well recognized by Hull with the use of his concept of behavioral oscillation ($_sO_r$) (3). Within the framework of Hull's theory, this concept would, in fact, lead to the prediction that for individual subjects, many inversions would be obtained in the determination of generalization gradients. The number of inversions which would be expected for any particular stimulus dimension would depend upon the range of the oscillation function, the form of the generalization gradient for that dimension, and the particular points on that dimension which were selected for test. The flatter the generalization function, the greater would be the expected frequency of inversions. It would also be predicted that there would be fewer inversions between stimuli which are far apart on the dimension than there would be between stimuli which are close together on the dimension. In general, this prediction seems to be verified by Razran's findings. The occurrence of inversions for individual subjects in generalization studies is almost directly analogous to the result that would be obtained from the psychophysical method of constant stimuli if only one or two judgments were obtained for each stimulus or stimulus pair. While a normal ogive may satisfactorily be fitted to data for the constant method if the number of judgments is large, a great many inversions are obtained if the number of judgments is small. Thus, while variability may obscure the lawful relationship which exists, it certainly is not used as an argument against its existence.

In addition to variability, another complicating factor enters in at this point to determine the nature of empirical generalization functions for individual subjects. This is the order in which the test stimuli are presented. This is an inherent difficulty in any generalization study in which the same subjects are tested for generalization at different points on the stimulus dimension. In another paper, the writer (1) has pointed out that, strictly speaking, only one valid measurement of generalization may be made for a particular subject. This is because of the possible effects of one test trial on subsequent ones. If test trials are reinforced, such reinforcement would have the effect of extending the generalization gradient. On the other hand, if test trials are not reinforced, the picture is complicated by the effects of differential reinforcement. Hovland's (2) solution to this problem was to counterbalance the order of presentation of the test stimuli. Thus, the generalization curve for any particular subject would be determined jointly by the generalization function and by the order in which the particular stimulus was presented. With this procedure, the fundamental law of generalization for this dimension could be determined only by determining the average response to each stimulus with order of presentation counterbalanced.

Similar reasoning with respect to variability should be applied to the argument concerning the lack of statistical significance between adjacent points when gradients are based on the results of groups of subjects. The obtaining of statistically significant differences depends upon the amount of variability, the size of the sample upon which the means are based, and the magnitude of the difference between the means. The magnitude of the difference between the means, in turn, depends upon the form of the generalization curve and the distance between the two stimuli on the stimulus dimension. Since the variability is usually relatively great and the size of the samples relatively small, it is not surprising that differences between adjacent or near points on the continuum are not statistically significant. The fact that statistically significant differences are found between points more widely separated on the continuum would appear to be proof that such gradients do, in fact, exist. The writer suggests that significance of trend tests based upon analysis of variance, such as the tests described by Linquist (4), would be a more appropriate type of statistical analysis to apply in this situation.

2. The second point to be discussed here is Razran's contention that generalization gradients are not generated at the time of original conditioning but that "all effects of generalization are generated during tests of generalization." Since the existence of generalization gradients can be known only by performing the operations necessary for their measurement, it would appear that it makes little difference whether the theorist assumes that the generalization gradient is generated at the time of original conditioning, at the time of the test, or at some point in between. Experiments involving additional training or changed

instructions subsequent to the original conditioning can have no bearing upon this question since such procedures might well be expected to modify the nature of the response in either case. Since generalization gradients may be determined only by actual test, statements to the effect that they are developed at the time of conditioning or at the time of testing appear to be operationally meaningless—at least insofar as behavior theory is concerned.

3. The final point to be considered is with respect to Razran's experiments on semantic generalization. While these experiments are of great importance and interest, they are not cases of primary stimulus generalization as that term is used by such writers as Hull and Spence. Semantic generalization may more properly be regarded as a form of secondary generalization or "mediated" generalization. It is quite likely that the mechanism underlying semantic generalization is quite different from the mechanism underlying primary stimulus generalization. While primary stimulus generalization gradients are assumed to depend upon physical similarity of the stimuli, it is quite likely that semantic generalization depends upon learning by the subject in which the conditioned stimuli and generalized stimuli have become associated. The postulation of some type of mediating process or response may be necessary to account for this phenomenon. While in no way minimizing the importance of mediated generalization, the writer would contend that a consideration of such phenomena is not necessarily relevant to the problem of primary stimulus generalization.

BIBLIOGRAPHY

1. GRICE, G. R., & SALTZ, E. The generalization of an instrumental response to stimuli varying in the size dimension. *J. exp. Psychol.*, 1950, **40**, 702–708.
2. HOVLAND, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, **17**, 125–148.
3. HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
4. LINDQUIST, E. F. Goodness of fit of trend curves and significance of trend differences. *Psychometrika*, 1947, **12**, 65–78.
5. RAZRAN, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, **46**, 337–365.

Received June 22, 1950.

REPLY TO GRICE'S "COMMENTS ON RAZRAN'S DISCUSSION OF STIMULUS GENERALIZATION"[1]

GREGORY RAZRAN

Queens College

At the time of the writing of my article I was familiar with Hull's *Principles of Behavior* and the postulate of behavioral oscillation, as well as with the text by Hilgard and Marquis and their concept of learned mediation as applied to the generalization from a stimulus object to its denoting word in an experiment by Smolenskaya (4, p. 201). I am grateful to Dr. Grice for the reminder but do not know in what way his comments call for any modification of my article and views. I shall take up Grice's three "points" in an order reversed from the one in which they are presented and will add two "points" of my own:

1. The statement that my experiments on semantic generalization are not unqualifiedly comparable with those of sensory (and spatial) generalization has my sympathy. But at no time did I say differently, as indeed I even questioned (after Lashley) the unqualified comparability of human sensory with animal sensory generalizations (6, p. 256; 7, p. 353). As the reader will remember, my chief source of evidence against the generalization views of Pavlov and Hull were results—numerous results—of Pavlov's own animal studies, and it is only when striking similarities between animal and human being and between sensory and semantic generalization had been unravelled that *some* common mechanisms were suggested. The obvious dependence of semantic generalization upon previous learning and association—to which Grice devotes half a dozen obvious sentences—hardly needed indication in an article whose task was to determine whether *all* generalization is not due to previous (pre-experimental) learning and association.

2. The argument (itself faulty; see 6, 8) of testable differences between generalization occurring during original conditioning and during subsequent testing is immaterial to my main thesis of suggesting a hypothesis that would, besides obviating a practically untenable neurophysiology, substitute present and on-the-spot and overt interactions and contiguities for past and remote and latent perseverations and potentialities. Should not a behavior theory, other things being equal, prefer that? And what about the "operational" advantages of having one less postulate to worry about?

3. The discussion (again quite obvious) of behavioral oscillation

---

[1] G. R. GRICE. Comments on Razran's discussion of stimulus generalization. *Psychol. Bull.*, 1951, 48, 150–152.

and of order of test stimuli presentation is equally of little relevance, since I do not deny lawfulness in CR generalization but merely claim the existence of a few crude categories rather than that of a continuum of j.n.d.'s or irradiations. And this is unopposed by what Grice says and, I repeat, will generally be admitted by unbiased readers of the total evidence, even though some readers might want in *some* instances more statistics to mediate between categories and continua, and others might put little—or insufficient or no—credence in my subjects' introspections. (I sometimes think that experimenters' ratiocinations do more harm—or less good—than subjects' introspections; after all, Descartes lived before Wundt.) Perhaps the word "bias" may serve as a clue to the differences between the thinking of Grice and myself. Grice, like a number of other indoctrinated behavior theorists, seems to attach a privileged status to Hull's postulates. To them these postulates are presumably true unless there is clear-cut evidence against them, while to me such postulates should not stay on the books unless there is clear-cut evidence for them (see Skinner for ambiguities of the term "postulate" in this connection, **9**). Nor do I always favor the dyadic dialectics of these postulates, even though it has been customary since the days of Zoroaster to make Ahriman, the god of darkness, shade the lines of Ohrmuzd, the god of light.

4. I recently analyzed a dozen more theses from Pavlov's laboratory as well as some more of my own data, and both suggest a supplementary hypothesis to the one of categorizing. It may well be that CR transfers that are now subsumed under primary stimulus generalization and under mediated generalization are really cases of residuals of multiple co-conditioning, as implied in Pavlov's studies of compound conditioning and as specifically stated by Hull (after Thorndike) as "generalization by means of identical stimulus components" (**5**, pp. 190–191). If a dog is conditioned to a metronome and a light and a tactile vibrator, and then shows CR transfer to a light and a tactile vibrator or to a light and a whistle and a tactile vibrator, there is little dispute about the mechanism of transfer. But it is obviously equally true that a dog conditioned to a certain tone frequency is also co-conditioned to the intensity, the duration, and the locus of the tone—besides of course other aspects in the CR situations—and when the tone frequency is changed, a good number of results of Pavlov's laboratory indicate that the CR transfer here is much like that obtained in the preceding case of compound conditioning. Again, in human subjects—as well as in some animal situations—the situation may be further modified through what Brogden called sensory pre-conditioning (**2**, **3**). Thus, if my subjects have been conditioned to the sound of a metronome (without seeing the metronome) or to the sight of the word "metronome," and then manifest CR transfer to the sight of the metronome (without sounding it), there is reason to believe—and subjects' introspections support the

belief—that the transfer is due to the fact that in the original CR situation the sight of the metronome was present and co-conditioned as a stimulus component through sensory pre-conditioning. Or, generalizing and reducing to lowest terms, we may say that in one case the conditioning is to "AB" and the generalization testing is to "AX" and that in the other case the conditioning is to "(A)B" and the generalization testing is also to "AX"—where "A," "B," and "X" are stimulus components (or aspects or even configurations) present during original conditioning or during subsequent testing, and "(A)" is a stimulus component invoked during original conditioning through co-conditioning—and thus obviate both primary stimulus generalization and mediated generalization.

5. One other point. Gradients of generalization in the manner of Pavlov and Hull contain, among other things, some unusual and contradictory assumptions about the capacity of organisms. Such gradients of generalization are obviously at the same time gradients of differentiation and thus assert that organisms are capable of developing in a few trials multitudes of discriminative reactions. Yet it usually requires a considerable amount of training to form even one discriminative reaction in the same organisms by the method of contrasts. And no human subject, I am sure, could emulate, with the best CR technique and voluntary effort, the reported gradient performance of any of Anrep's three dogs (1). I defy any psychologist to do so.

BIBLIOGRAPHY

1. Anrep, G. W. Irradiation of conditioned reflexes. *Proc. roy. Soc.*, 1923, 94B, 404–426.
2. Brogden, W. J. Sensory pre-conditioning. *J. exp. Psychol.*, 1939, **25**, 323–332.
3. Brogden, W. J. Human sensory pre-conditioning. *J. exp. Psychol.*, 1947, **37**, 527–539.
4. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning*, New York: Appleton-Century, 1940.
5. Hull, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
6. Razran, G. Some psychological factors in the generalization of salivary conditioning to verbal stimuli. *Amer. J. Psychol.*, 1949, **62**, 247–256.
7. Razran, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, **46**, 337–365.
8. Razran, G. Semantic and phonetographic generalizations of salivary conditioning to verbal stimuli. *J. exp. Psychol.*, 1949, **39**, 642–652.
9. Skinner, B. F. Review of Hull's *Principles of behavior*. *Amer. J. Psychol.*, 1944, 57, 276–281.

Received July 14, 1950.

A STATISTICAL CONSIDERATION IN PSYCHOLOGICAL RESEARCH

BRYAN WILKINSON

Prudential Insurance Company, W.H.O.

In much psychological investigation, the experimenter is confronted with several variables. He hopes that certain significant relationships, or differences, may exist between various pairs of these variables without having logical grounds, a priori, for stating just which pairs they might be. To test his idea, the investigator may calculate all possible correlation coefficients, critical ratios, *t*-ratios, *F*-ratios, or chi-squares—depending upon the experimental design. Those pairs of variables which show a significant relationship, or difference, are then investigated further—or the experiment may stop at this point with conclusions based upon the "significant" findings.

While the statistically rigorous investigator may find some quarrel with the above procedure (1), frequently it is almost the only avenue of investigation open. This is particularly so where the subjects are not rats or sophomore psychology students but are instead adults in the clinic or industry.

Assuming, then, that the above procedure has been followed, there results a mass of data which must be analyzed. In this analysis, certain principles regarding significance of findings must be observed (3). In the simple experiment where one correlation, *F*-ratio, *t*-ratio, or chi-square is calculated, the experimenter selects the 5 per cent, 1 per cent, 0.1 per cent, or some such level of significance and decides that his results are or are not significant. In the "blind analysis" type of experiment the procedure is not so simple. Using correlations as an example, were 100 *r*'s to be computed, about five of them would differ significantly from zero at the 5 per cent significance level even though there was actually no correlation between any of the pairs of variables—this by definition of the 5 per cent significance level. Thus, the experimenter who finds two *r*'s, out of the 14 he has calculated, significant at the 5 per cent level may not have two events each one of which would happen only five times in 100 trials. Instead, he probably has one event which would happen about 15 times in 100 trials on the basis of chance alone. This probability was based on the binomial expansion (2).

Whenever a large group of statistics is collected in the "blind analysis" type of experiment, the number of significant statistics should be tested by the binomial expansion to determine if that number is sufficiently great to indicate a non-chance occurrence. (This applies to groups of correlation coefficients, *t*-ratios, critical ratios, *F*-ratios, and chi-squares.)

Of course, expanding the binomial to determine the probability involved can be a laborious task. Although certain tables are available

TABLE I

PROBABILITY OF OBTAINING *n* OR MORE SIGNIFICANT STATISTICS BY CHANCE IN A GROUP OF *N*

(Where statistics are significant at 1 per cent level.)

n	1	2	3	4
N				
1	.0100			
2	.0199	.0001		
3	.0297	.0003		
4	.0394	.0006		
5	.0490	.0010		
6	.0585	.0015		
7	.0679	.0020		
8	.0773	.0027	.0001	
9	.0865	.0034	.0001	
10	.0956	.0043	.0001	
11	.1046	.0052	.0002	
12	.1136	.0062	.0002	
13	.1225	.0072	.0003	
14	.1313	.0084	.0003	
15	.1399	.0096	.0004	
16	.1485	.0109	.0005	
17	.1571	.0123	.0006	
18	.1655	.0138	.0007	
19	.1738	.0153	.0009	
20	.1821	.0169	.0010	
21	.1903	.0185	.0012	.0001
22	.1984	.0189	.0013	.0001
23	.2064	.0220	.0015	.0001
24	.2143	.0239	.0017	.0001
25	.2222	.0258	.0019	.0001

which expand $(p+q)^n$ where $p=.05$ or .01, it is the purpose of this paper to present tables which will furnish the psychologist with the probabilities of obtaining *n* or more significant statistics in a set of *N* on the basis of chance alone. It is to be understood that these tables should not necessarily be referred to where a priori, rather than a posteriori, logic

TABLE II

PROBABILITY OF OBTAINING *n* OR MORE SIGNIFICANT STATISTICS BY CHANCE IN A GROUP OF *N*

(Where statistics are significant at 5 per cent level.)

n	1	2	3	4	5	6	7
N							
1	.0500						
2	.0975	.0025					
3	.1426	.0072	.0001				
4	.1855	.0140	.0005				
5	.2262	.0226	.0012				
6	.2649	.0328	.0022	.0001			
7	.3017	.0444	.0038	.0002			
8	.3366	.0573	.0058	.0004			
9	.3698	.0712	.0084	.0006			
10	.4013	.0861	.0115	.0010	.0001		
11	.4312	.1019	.0152	.0016	.0001		
12	.4596	.1184	.0196	.0022	.0002		
13	.4867	.1354	.0245	.0031	.0003		
14	.5123	.1530	.0300	.0042	.0004		
15	.5367	.1709	.0362	.0055	.0006	.0001	
16	.5599	.1892	.0429	.0070	.0009	.0001	
17	.5819	.2078	.0502	.0088	.0012	.0001	
18	.6028	.2265	.0581	.0109	.0015	.0002	
19	.6226	.2453	.0665	.0132	.0020	.0002	
20	.6415	.2642	.0755	.0159	.0026	.0003	
21	.6594	.2830	.0849	.0189	.0032	.0004	
22	.6765	.3018	.0948	.0222	.0040	.0006	.0001
23	.6926	.3206	.1052	.0258	.0049	.0008	.0001
24	.7080	.3392	.1159	.0298	.0060	.0010	.0001
25	.7226	.3576	.1271	.0341	.0072	.0012	.0002

or the experimental design points to several pairs of variables which should be significantly similar, or different.

BIBLIOGRAPHY

1. FISHER, R. A. *The design of experiments*. Edinburgh: Oliver & Boyd, 1935.
2. HOEL, P. G. *Introduction to mathematical statistics*. New York: John Wiley, 1947.
3. MCNEMAR, QUINN. *Psychological statistics*. New York: John Wiley, 1949.

Received June 16, 1950.

BOOK REVIEWS

STONE, C. P. (Ed.) *Annual review of psychology.* Stanford, Calif.: Annual Reviews, Inc., 1950. Pp. vii+330. $6.00.

The recent increase in the quantity of psychological publications, as well as the fission of psychology into many diverse areas, has prevented most psychologists from maintaining an adequate panoramic view of the development of their science. Because of this condition, there is urgent need for some sort of periodic review of contemporary psychological literature which will enable specialists in one branch of the subject to keep in touch with other areas. The first volume of the *Annual Review of Psychology* suggests this need will be competently met.

This volume contains reviews of recent psychological literature (1948–1949) in 18 fields of psychology. The aims set forth by the Editorial Committee were twofold: (1) that the reviewers "cover principally publications . . . which [they] considered most noteworthy. . . .," and (2) that the reviewers "adopt an interpretative and evaluative approach to the literature they selected for review." What is meant by "noteworthy" obviously depends upon individual judgment as to what is the most appropriate research strategy to pursue at the present stage of development in the science of psychology. It therefore seems justifiable to demand of each reviewer an explicit statement of his rationale for the inclusion and exclusion of publications. Many of the authors supply such a statement, but most do not.

Those readers who are familiar with the Robinson-McGeoch-Melton tradition in the psychology of learning will be surprised to discover that Arthur W. Melton, in his top-notch chapter on recent learning publications, chooses to organize it "in terms of the major theoretical problems of the learning process, including transfer of learning and retention, but with emphasis on the theoretical differences of the Tolman and Hull groups." This writer would be the last to question the wisdom of the selection of such an organizing principle, but it is quite possible at our present primitive stage of development that important findings need not be instigated by *theoretical differences.* This comment in no way detracts from the general excellence of Melton's chapter.

The method by which the various writers chose to deal with the literature under review varied over a wide range. Robert L. Thorndike, in one of the best chapters, "Individual Differences," does much more than merely report the findings of the numerous studies. He comes to grips with the ideas behind the studies and does not hesitate to be critical when necessary. Robert R. Sears, "Personality," and Carroll L. Shartle, "Industrial Psychology," in two other good chapters, follow somewhat the same procedure. Because Sears reviews only 39 publications, he is able to deal more intensively with the literature he cites. One wonders, however, why the field of personality, which has captured the interest of such a large number of psychologists, has the shortest list of references.

The field of discriminal processes is handled expertly by Neil R. Bartlett, "Vision," Edwin B. Newman, "Hearing," and Frank A. Geldard, "Somesthesis and the Chemical Senses." It is interesting to note Bartlett's concern about the neglect of some important visual problems due to the tendency of Federal agencies to support research most likely to have practical application. Newman's chapter attempts, unlike most other reviews, to integrate apparently discordant data. His scholarship is impressive, as is his pleasant prose; the latter, unfortunately, appears in this volume much too infrequently.

All of the reviews are adequate and informative. Some are particularly interesting because they reflect so well the existing state of the field which is reviewed. Jerome S. Bruner's chapter mirrors the disorganization of present-day social psychology. It would have been helpful if Bruner had attempted to relate more explicitly the research programs of social psychologists to their underlying assumptions as to the nature and aim of science. Perhaps then a somewhat clearer picture of contemporary social psychology would have emerged.

The chapters in the clinical areas would benefit by the elimination of overlapping. There are two chapters on therapy ("Clinical Methods: Psychotherapy" and "Counseling Methods: Therapy"), as well as two chapters on diagnostics ("Clinical Methods: Psychodiagnostics" and "Counseling Methods: Diagnostics"). It would be wise for future volumes to compress these four chapters into two, one on psychotherapy, the other on psychodiagnostics. Howard F. Hunt's chapter on psychodiagnostics and William V. Snyder's review of psychotherapy are both excellent.

David A. Grant's treatment of "Statistical Theory and Research Design" is very complete but, unfortunately, pitched at a rather high level. Most psychologists must keep up with new statistical techniques, and it would have been quite helpful if the review of statistical procedures had somehow been designed to meet their needs.

The editors were no doubt pressed to meet a deadline. Several errors in the lists of references testify to this. It would be helpful if the future volumes in this series presented the bibliographies in alphabetical order. Hunt's chapter, in the present volume, is the only one following this procedure.

One cannot read this book in its entirety without reflecting upon the expansiveness of contemporary psychology and its concomitant lack of cohesiveness. To some extent the book intensifies the confused feeling common to so many psychologists attending APA meetings who seek a rationale for the existence of one organization representing all of psychology. Although not convinced that his opinion as to the mutual interdependence of the different levels of psychology is justified—either on methodological or on practical grounds—this reviewer nevertheless feels that it would be most unfortunate if the definite trends toward fission in psychology were to continue at their present pace. Do not psychologists in all areas have common problems? Can we not assist each other in the solution of some of these problems? The frequent "ethnocentric" complaints heard from all the "sub-cultures" of our psychological community stem mainly from lack of information. If all psychologists read the *Annual Review of Psychology* in its entirety—as this writer believes they should—then better understanding might develop between workers in different fields of psychology. Perhaps here will lie the major contribution of this volume.

Howard H. Kendler.

New York University.

Sargent, S. S. *Social psychology: an integrative interpretation.* New York: Ronald Press Co., 1950. Pp. x+519. $4.50.

Sargent's book adds another volume to the spate of competent social psychology texts that have appeared since the war. Eclectic and relatively elementary, it will recommend itself for its broad coverage, up-to-dateness, and general good sense rather than for theoretical incisiveness or detailed treatment of important research areas. Greater emphasis is given to sociological contributions than is usual in a psychologist's social psychology; indeed, most of the present and traditional preoccupations of social psychologists of all persuasions come in for some attention. The result has both the virtues and the limitations of an informed eclecticism.

Part I treats the socialization of the individual, taking as a starting point the heredity-environment issue in social psychology and proceeding, via a consideration of anthropological and sociological evidence concerning social influences on personality, to an elementary discussion of the process and products of social learning. Part II is devoted to the various theories of social motivation, with primary attention to psychoanalytic mechanisms and to Allport-Cantril-Sherif ego theory. In Part III, the "patterning of social behavior" is considered in regard to language and communication, the forms of social interaction, social groups, social roles, and leadership. The section ends with the presentation of a simple "field theoretical" orientation in which social behavior is attributed to three broad categories of interrelated determinants: personality, the social situation, and the individual's perception of the situation. Applications of social psychology to the understanding of social phenomena comprise Part IV, which completes the book. The chapter headings indicate the miscellany of topics singled out here for treatment: public opinion, propaganda, mass behavior, social change and social movements, group differences and prejudice, and applications in the area of social issues. A final summary chapter attempts a judicious appraisal of pure and applied trends in current social psychology.

The chapter on social roles adds most toward clarifying present terminology and assumptions. Yet Sargent's proposed definition seems not to avoid entirely the existing confusions that he depicts. Since the patterned social expectations concerning a person's behavior, the situational patterning in his behavior itself, and situationally organized facets of his self-image may vary independently, it is hardly satisfactory to embrace them as he does under a single definition. Also noteworthy are the particularly clear-headed analysis of hereditary and environmental influences, and the brief but comprehensive survey of action research associated with the Society for the Psychological Study of Social Issues.

Certain gaps stand out in the generally full coverage. The extensive recent research on the social psychology of the cognitive processes is merely mentioned in passing. The California studies of personality dynamics in prejudice are ignored, except for a notation on supplementary reading. Research methods, discussed briefly at the outset, are given rather little attention thereafter, and the interpretative conclusions of studies are typically cited without much detail as to the data on which they were based. Supplementary material is clearly called for if this text is to be employed in courses with a research emphasis.

A degree of superficiality is the necessary cost of attempting so broad a scope. More serious as a limitation is the loose integration characteristic of Sargent's eclectic approach. Conceptually, the lion is sometimes made to lie down with the lamb, disregarding possible incompatibilities, as when Miller and Dollard terminology is coupled with what is essentially a cognitive field theory. There is rather little theoretical interpenetration among the major sections of the book. Roles, attitudes, and socialization are discussed in different contexts, and the treatment of applications is little tied to the earlier theoretical sections. Conventional categories like imitation and suggestion, whose usefulness has recently been challenged, are preserved intact, and the book is replete with current classificatory schemes of somewhat dubious usefulness.

References to the literature, given in footnotes, are extensive and up to date. A helpful teaching aid is provided in references at the end of each chapter to the relevant sections of Newcomb and Hartley's *Readings in Social Psychology* and Britt's *Selected Readings in Social Psychology*, as well as to more intensive sources. The presentation is clear and well organized, and the frequent recapitulations and summaries should prove particularly helpful to the student.

Vassar College. M. BREWSTER SMITH.

GURVITCH, GEORGES. (Ed.) *Sociometry in France and the United States. A symposium.* New York: Beacon House, 1950. Pp. vii+261. $7.50.

The first three numbers of Volume 12 of the journal *Sociometry* have been bound in book form to make an aniversary volume. Roughly the sixteen articles constituting the book can be divided into three types: seven contributions of a speculative, theoretical or critical nature dealing with sociometry, microsociology and axiology; three articles giving the history, application, and present centers of sociometric work; and six reports of research in which sociometric techniques have been employed.

Much of the theoretical and speculative discussion is uninhibited in its praise of the sociometric techniques. J. L. Moreno unabashedly takes credit for saving all the social sciences from sterility and also for establishing the basis for a dynamic social revolution. "With the cooperation of 'all' the people we should be able to create a social order worthy of the highest aspirations of all times. This, and this alone is the meaning of revolutionary, dynamic sociometry" (p. 254). His critical analysis of Marxism and his "sociometric theses" of social change are of real dialectic interest; but their experimental and practical soundness are, in the reviewer's opinion, yet to be demonstrated. Nevertheless, Moreno does make a cogent analysis of some of the defects of much modern sociological and social psychological research.

Georges Gurvitch, in an article entitled "Microsociology and Sociometry," presents a detailed description of his concept of microsociology. In general, Gurvitch is favorably inclined toward the sociometric approach but laments the extent to which the "We-ness" of social relations is untapped by sociometry. René Zazzo considers the relations between sociometry and psychology. While recognizing the value and uniqueness of Moreno's work, Zazzo believes that it overemphasizes the affective nature of interpersonal relations. Leopold von Wiese discusses sociometry as represented in Moreno's *Who Shall Survive?* Paul E. Johnson presents a completely speculative neotheology. It remains for Pitirim Sorokin to examine critically Moreno's concepts of spontaneity and creativity. He raises a number of critical questions such as the extent to which creativity and spontaneity, as measured by Moreno's techniques, are reflctions of *genuine* spontaneity and creativity, and not simply manifestations of an uninhibited superficiality. In a final article Moreno attempts to answer these criticisms of sociometry.

Helen Jennings gives an historical account of the military use of sociometric and situational tests in a number of countries. She praises the Office of Strategic Services program for retaining a large element of spontaneity in its assessments. A listing of sociometric research and work centers in France is given; and Zerka Toeman presents a "History of the Sociometric Movement in Headlines."

The research reports are of great interest both because they tend to give an operational definition of one aspect of sociometric work and because of the results they present. Paul Maucorps reports on an intensive sociometric investigation of thirty-five officers being trained in a French military school for staff work. Bengt Danielsson's study of an Amazon Indian group suggests an operational definition of "key persons" in simple groups and also presents some positive evidence of the relationship between verbal and nonverbal behavior. Charles Loomis and Reed Powell report a study of the different types of class status in a Costa Rican peasant proprietor area and in a hacienda community. The lack of communication between social classes in the centralized rural plantation, in contrast to the mobility of the proprietor group, is noteworthy. George Lundberg and his co-workers present the results of a study of the attraction and repulsion patterns in a university. Apparently propinquity is the major determining factor in these patterns. Henry Maas reports a not altogether successful attempt to relate changes in social attitude to changes in sociometric choice. Finally the excellent study of buddy ratings by Robert Wherry and Douglas Fryer is reprinted from *Personnel Psychology*. The authors were able to gather a number of criteria of leadership at the Signal Corps Officer Candidate School and, by careful analysis, show that anonymous nominations constitute their purest measure of leadership.

For those interested in social psychology and who do not have the journal *Sociometry* available, this book offers a stimulating review of this growing field. The reviewer cannot help but add that the book contains many typographical errors and parts of some charts are illegible; the price, $7.50, seems excessive.

University of Rochester. LAUNOR F. CARTER.

ROBACK, A. A. *Personality in theory and practice.*[1] Cambridge, Mass.: Sci-Art Publishers, 1950. Pp. 437. $3.75.

This is another original and thought-provoking work by the learned author of the monumental treatise, *The Psychology of Character*. In his characteristic style he leads us with an almost deceptive facility through the maze of the scholarly and scientific contributions to the multi-chambered castle of personality and then offers his own definition: "Personality is the integrated organization of all the cognitive, affective, conative, and physical characteristics of an individual as it manifests itself in focal distinctness *and carrying a special meaning* to others" (the italicized words refer to the addition to his original definition in Warren's *Dictionary of Psychology*). This latter phrase has been added in accordance with his view that personality, like character, should be evaluated. In this respect, then, Roback's view is much like May's concept of "social-stimulus value," and likewise has the weaknesses which have been noted by several writers, notably, G. W. Allport. One might also criticize the word "integrated" in Roback's definition, connoting as it does *un fait accompli*, by asking to what *degree* integration must be achieved in order that the term "personality" may be applicable. Are children, for example, automatically disfranchised as possessors of personality because their "integrations" are unfinished?

---

[1] The original book, now out of print, *Personality, the crux of social intercourse* (1931), constitutes about a fifth of the present text. A review of it by G. W. Allport appeared in *Psychol. Bull.*, 1932, 26, 359–360.

After a discussion of "trait" as the unit of personality which follows along the lines drawn by G. W. Allport, we are treated to a challenging chapter on the determinants of personality. Here Roback seeks to lay bare the implicit assumptions of the strictly environmentalistic interpretations. But he is not so naïve as to ignore completely the effects of environment. Thus, he writes, "our dispute is not with a moderate geneticism, but with an immoderate environmentalism." Undoubtedly psychologists—perhaps, more especially American—under the combined impact of behaviorism, sociology, and cultural anthropology, have been neglecting, or worse still, ignoring the nativistic determinants of personality. But, ultimately, we must make a choice between these two contenders. Shall the choice be on the basis of some aesthetic or metaphysical quality inherent in the one or the other, or on their respective merits to initiate crucial experiments or empirical studies to prove or disprove themselves? The latter alternative, of course, is more in keeping with our scientific interest and should become a competitive motive between the two camps. Roback's uncompromising but witty polemic sets well the stage for the first round in the comeback of the undying instinctivist position.[2]

Perhaps the most original ideas in the book are to be found in the chapters on "Meaning and the Personal Idiom," "The Adequacy Valence of Personality," and "The Criterion of Personality." In the first of these Roback insists—not unlike the Freudians—that our seemingly meaningless acts which, at first blush, appear as *disjecta membra*, have, nevertheless, a meaning peculiar to ourselves—our "personal idiom." Furthermore, were it not for the differences in these acts from person to person we could scarcely be aware of the "aroma" of personality. Good and well. But how shall we ascribe meaning, and from whose frame of reference? Here, however, we are left to our own devices. In the following chapter Roback defines his concept, "adequacy valence," as "a personality nucleus which . . . represents an ideal relationship

---

[2] Another recent provoking treatise written from the same vantage point is Helge Lundholm's *God's failure or man's folly?*, Cambridge, Mass.: Sci.-Art Publishers, 1950.

Bureau Ednl. Psy. Research

between the individual and society." Thus defined, he then seeks out "the ingredients of our personality which together constitute human worth. . . . " He finds it in a triad: independence, insight, and sentiment. This selection, of course, is arbitrary and suffers accordingly the fate of all such selections in ethics as elsewhere. Nevertheless, the idea underlying Roback's selection is commendable. Psychologists measure a person's intelligence, attitudes, aptitudes—why not his total worth? In the chapter, "The Criterion of Personality," the author denies that personality evaluated is coterminous with character. Rather, the latter "connotes a fighting spirit on behalf of the regulative principles," while the former connotes sympathy and "the extent of the influence a personality exercises for the general good over a period of time." It is hard to see the boundaries that Roback tries to erect here, for the "regulative principles" (liberty, justice, truth, world progress, etc.) seem to mean the same thing as "the general good." Nevertheless, people do evaluate one another as "good," "bad," etc., and psychologists, armed with refined techniques, might well profit by a concern with these worthy matters. Whether they call it ultimately evaluation of character or evaluation of personality should not matter. What matters is evaluation!

In fine, this book is like a gadfly dissatisfied with much of present day psychology. Its needling questions ought to go far towards disturbing psychologists out of their complacency about "accepted" answers and prodding them into a rethinking of their research ends and research means. This book probably will not be acceptable as a college text to most psychologists of personality, but it might well prove to be of even greater usefulness if they themselves became its students. Indeed, no one interested in psychology should miss reading it.

Institute of Child Welfare, University of Minnesota. PERCY BLACK.

MAY, ROLLO. *The meaning of anxiety.* New York: Ronald Press Co., 1950. Pp. xvi+376. $4.50.

This is a book which paints the problem of anxiety on a wide canvas. The author sees anxiety as "a pervasive and profound phenomenon" of our time, a "nameless and formless uneasiness that has dogged the footsteps of modern man." Today it has become a conscious and explicit problem, recognized as central and urgent not only by students of neurosis but also by poets and philosophers, historians and social scientists, teachers and ministers, in fact by all who seriously concern themselves with human affairs. Heightened by the grim external circumstances of the immediate present, anxiety is deeply rooted in those rapid cultural changes which produce "inner confusion, psychological disorientation, and uncertainty with respect to values and acceptable standards." We live in an age of anxiety and we need to be as clear as we can about it. To this end the author brings together the theories of anxiety "offered by modern explorers in different areas of our culture," searches them for common elements, and attempts to construct a coherent and comprehensive synthesis. Though written by a psychologist, the book is not intended only for professional colleagues but for all serious students of human problems.

The generous scope is apparent throughout the first half of the book. From the philosophical predecessors of modern theorists of anxiety the author chooses Spinoza, Pascal, and especially the subtle Kierkegaard. His biological sources include Goldstein, Cannon, and the current workers in psychosomatic medicine. Animal and child psychology are searched, then psychotherapy, an order of presentation which gives one of the odd sensation of jumping from Mowrer back to Freud but which valuably widens the now well worn path from Freud through Adler to Horney and Sullivan. The cultural approach is chiefly represented by Tawney, Fromm, and Kardiner, but wide historical reading also informs this stimulating chapter. There is a refreshing freedom from departmentalized thinking; rather, the progress through this series of skilfully condensed surveys has the character of a continuous journey which places the problem of anxiety in the full perspective of human life.

The first and larger section of the book is brought to a close with a chapter entitled "Summary and Synthesis of Theories of Anxiety." It is more summary than synthesis. My first feeling, however, that the author had failed to achieve a tight and logical synthesis was soon tempered by the realization that he had perhaps hoped for more synthesis than the nature of the problem allowed. Part of the problem of anxiety is to discover the nature of the reaction itself and to work out both the gross and subtle effects on personality produced by the necessity for defense. Another large part, it seems to me, consists of pointing out the myriad circumstances in which anxiety is aroused—myriad because personality is essentially complicated, ramified, set in a host of physical and cultural circumstances. Freud, Horney, Kierkegaard, Fromm point out somewhat different circumstances under which threat arises; all are "true" and they do not have to be synthesized. My only serious criticism of this chapter, therefore, is of the author's distinction between anxiety and fear, a point to which he devotes considerable discussion. He makes the distinction by means of two criteria: (1) anxiety is diffuse and objectless, fear specific and object-related; (2) anxiety is provoked by threat to some value felt essential to existence as a personality, fear is provoked by threat to some merely peripheral value. I do not believe that these two criteria can be worked simultaneously, and in any event I wonder if the distinction between anxiety and fear is not more verbal than biological.

The second part of the book consists of case studies, one from the author's psychoanalytic practice and thirteen from a special study of unmarried mothers at a refuge home. In reading these cases I had a sense of shrinkage from the wide scope of the rest of the book to the constricted dimensions of the consulting room. I did not feel that the reports were combed free of Rorschach clichés and pat psychodynamic formulations, especially those concerning rejection by the mother, and I questioned whether the case material was always adequate to support the conclusions drawn. The cases, to be sure, are not offered in proof of anything; rather, they are intended to illustrate different sources of anxiety and personal patterns of dealing with it, and they have the merit of displaying normal and successful adjustments as well as neurotic ones. Yet even when they are accepted in this light they seemed to me to belong only at the common level of case reporting, not on the high ground that is occupied in the main part of the book.

Harvard University. ROBERT W. WHITE.

HARSH, C. M., AND SCHRICKEL, H. G. *Personality development and assessment.* New York: Ronald Press Co., 1950. Pp. vii+518. $5.00.

Books on "personality" have many different aims and contents. This recent textbook is intended for a survey or coverage of a wide variety of studies and theories. It aims at being introductory but thorough and to present material in a way that will "minimize theoretical biases . . . [and] . . . aid development of an inclusive and unified perspective toward personality study." The book is organized into two main parts. The first part deals with the nature and development of personality and is organized into chronological periods beginning with infancy and ending with old age. Most of this section (296 pages) deals with generalizations made from experimental investigations and affords an excellent source and organization of a great many studies which might be classified as child development, personality theory, and studies of human learning. This section also includes many specific hypotheses gleaned from clinical observations and theoretical deductions.

The first three chapters in Part II deal with different theoretical orientations to personality. These chapters serve to acquaint the student with some of the broad trends in personality theories, but they do not present a systematic or thorough comparison of personality theories. Compared to the developmental survey portion of the book, this is a less useful section. Although the authors, in general, draw on learning experiments for much of their material, perhaps a major omission here is to fail to give some attention to the beginnings of attempts to approach personality from a "learning theory" point of view as a new but still poorly systematized approach. The remaining chapters in Part II deal with an introduction to personality measurement or assessment and the problem of trait organization.

The authors would be hard pressed to justify as unbiased the selection of concepts used, the amount of space devoted to various topics, and the inclusion and exclusion of topics. Surely there is some implicit theory or bias to this book, as in all books, and perhaps it would have been better to make it explicit and then have the opportunity not only to organize data into logical headings but also into systematic constructs. Failure to do this involves the necessity to cover many overlapping concepts (i.e., anxiety, inferiority, guilt, conflict, etc.) as if they were all quite different things. Although the authors are aware of the problem, this book perhaps abets rather than discourages the proliferation of non-operationally defined constructs in this area where investigation and prediction are already relatively difficult.

This book remains, however, a highly readable and informative book which will serve as an excellent text in beginning courses in personality and development and which will be enjoyed by students. It might be well for instructors to supplement the text with some discussion of the need for operational definition and systematic formulation of experimentally testable hypotheses in a field where communication is low and confusion high.

JULIAN B. ROTTER.

The Ohio State University.

NEWCOMB, THEODORE M. *Social psychology.* New York: Dryden, 1950. Pp. xi+690. $4.50.

This newcomer among social psychology texts consummates the wedding of psychology and sociology in the author's thinking, research, and teaching. Newcomb has been a leader among those who are convinced that psychology and sociology complement one another and who have been striving to bring these disciplines into more intimate relations. His *Social Psychology* should enhance his leadership in this inter-disciplinary group since it appears to do a distinctly better job of relating certain key concepts and principles of the two fields than any previous text.

An introductory text in social psychology should offer the student a synthesis of the field in terms of concepts and principles which reflect a consistent theoretical viewpoint. Newcomb's text embodies this virtue. Social psychology is defined as the study of individuals interacting, that is, perceiving and responding to each other's responses. Sociology (or cultural anthropology) enters the picture inasmuch as individuals usually interact in structured group contexts where they are under constraint to perform their prescribed *roles* in compliance with group *norms*. The bulk of this text is devoted to showing how (or how profoundly) group roles and norms affect the individual's personality and his interaction with others. To this end, Newcomb marshals psychological, sociological, and anthropological research findings with such vigor and emphasis that the student can hardly fail to accept his central thesis. Students will probably be especially impressed by his use of anthropological data to relate culture and personality.

Throughout the book, the stress is on the individual (albeit as group member), and the theory presented is largely psychological theory. Much of Newcomb's theory is drawn from general psychology, in particular the terms and principles of motivation, learning, and perception. Adding the adjective "social" to these terms does not call for new theoretical formulae; rather, the same formulae are now filled with items relevant to social interaction instead of such items as maze pathways and pathway choices. Other important segments of Newcomb's approach derive from personality theory, notably his consideration of personality conflict and ego psychology. Of particular interest is his well-reasoned account of the ways in which Freud's and Horney's theories contribute to an understanding of the socialized personality.

Newcomb makes socialization the primary focus of his analysis. Socialization is described as a process in which the initiate learns to satisfy his *drives* or *motives* (organic and socially acquired) by developing *motive patterns* and *attitudes* which correspond to the roles and norms of the groups in which he holds membership. A key term throughout the book is *motive pattern*, which is defined as an organization of the perceptive, cognitive, affective, and performance aspects of one's responses toward a particular goal. In other words, perceiving, thinking, feeling, and doing are viewed as significant in relation to the motives which organize them and which they serve. An *attitude* is defined as a predisposition to have a motive toward its object, and *role patterns* are viewed as organized in terms of the motives they express.

While social motives, motive patterns, attitudes, frames of reference, and role patterns are discussed as learned, Newcomb nowhere offers a detailed analysis of the social learning process. Thus, he uses six pages to discuss how motives are acquired (121–127). He reminds the reader of the trial-and-error and conditioning formulations and posits the acquisition of second-order drives as bases for elaborating motive patterns. For further details of the learning process, he refers the reader to other sources rather than himself testing the applicability of various learning principles to social motivation. Learning new motive patterns is discussed as an aspect of personality formation. In this account, new motive patterns are said to develop when others are blocked. Thus motives will out, though little is said about how new channels of expression are selected.

In outlining the processes of attitude formation, Newcomb uses four pages to assert that attitudes develop out of specific acquired drives, become differentiated, and organize into inclusive hierarchical systems (127–131). He justifies this scant development of the topic by pointing out how sparse are the research findings on attitude formation in childhood. Attitude change, on the other hand, is fairly well outlined by research findings, and Newcomb devotes the bulk of two chapters to this subject. This account is largely in terms of perceptual reorganization rather than learning theory, with emphasis placed on changes in socially conditioned frames of reference.

One of the more difficult decisions confronting the writer of a social psychology text is how fully to discuss personality structure. Newcomb centers his discussion on those aspects of personality which are most evidently affected by social influences. He recognizes the role of constitutional factors in personality organization and attributes "temperamental" or "expressive" traits to them. Legitimately, he elects not to discuss these factors, limiting his account to "directive" factors such as motives, motive patterns, self-other attitudes, and frames of reference, all of which are discussed primarily in their relations to group influences. He devotes considerable attention to personality dynamics, particularly in relation to the concept of "self," the hierarchical structuring of attitudes, and the interplay of motives, perceptions, and attitudes.

Individual similarities and differences in personality as revealed in attitudes and social behavior offer an important test of Newcomb's account of personality in terms of group influences. He passes this test well by accounting for similarities in terms of common group experiences, differences in terms of distinctive group experiences. The uniqueness of individual personality is easily accounted for by pointing to the unique combinations of social influences with which each individual's multiple group memberships supply him. An important contribution of Newcomb's account of personality in society is his discussion of ways in which personality conflict is related to conflicting roles and norms in the various groups to which the individual belongs.

If one looks for the contributions of sociological theory to Newcomb's viewpoint, he finds essentially an analysis of group structure in terms of interlocking roles and group norms (Chap. 8). It is hard to find a page in the book where the discussion does not relate to roles or norms. However, Newcomb makes clear that these group factors affect individuals only through interaction with other individuals. And so the theory called for in social psychology is primarily a theory of the interaction of individuals in face-to-face groups. As Newcomb points out, the interaction of individuals is the meeting-ground of individual psychology and sociology. Appropriately, sociologists (such as Bales with his "interaction process analysis") and psychologists are now working jointly toward developing interaction theory.

To this reviewer, the greatest missed opportunity in Newcomb's text is his failure to venture a specific and integrated account of interaction. This may seem an odd complaint since he builds his entire presentation around analyzing the effects on the individual of his interaction with others, as well as discussing ways in which the individual may influence the groups in which he performs roles. However, Newcomb rejects the task of classifying and analyzing various forms of interaction such as helping, cooperating, attacking, and loving. Rather, he concludes that "the major social-psychological problem . . . is not that of listing and classifying the possible forms but of discovering the *conditions* under which an identifiable kind of interaction does or does not take place" (p. 24). By and large, it seems fair to say that Newcomb deals more with the effects of group roles and norms on individuals than with the processes of interaction which accomplish those effects.

What sort of interaction theory is now proposed? This reviewer's suggestion is that interaction theory be developed on the premise that the simplest and most fundamental instances of interaction involve two individuals (the "dyadic group"). Viewed in their group context, the two would be seen as performing mutual roles which defined the modes of their interaction. Each would be viewed as bringing to the interaction his own motives, attitudes, and frames of reference. Each segment (episode) of their interaction would be interpreted or evaluated in terms of how it affected their perceptions, attitudes, and motives and how it affected the course and products of their interaction. Presumably, the same theory would apply to larger groups, though additional terms and principles would probably be needed to account for the sub-groupings of members. Newcomb has certainly pointed to the need for such a theory as the one proposed, and has offered many valuable leads toward its development. It is a matter for regret that he has not assembled his ideas into a more systematic account of interaction.

Newcomb's text strongly represents the scientific approach. He defines his terms operationally and points to the necessity for intervening variables, such as motive, which must be measured indirectly. He stresses theory and makes excellent use of research findings to illustrate or test the hypotheses he advances. He shows the utility of his theory for working out solutions to practical problems such as overcoming prejudices and reducing group conflicts.

The text would have been strengthened by a discussion of observational methods and by a consideration of problems encountered when designing and conducting social-psychological research in the laboratory and in "real-life" situations. Some discussion of the relations between basic and applied research would have helped the student understand the significance of current research trends in the field.

Viewed as literature, this text is a superior job of exposition. Illustrative materials, drawn largely from research studies, are well chosen for scope, relevance, and interest value. Each chapter begins with a section calculated to evoke the student's interest, followed by a section stating the chapter objectives. Summaries are frequent, and highly effective use is made of parenthetic remarks to answer on-the-spot questions which students are apt to raise. Probably the tables would be improved by supplying titles, and more consistent use might be made of italics in defining terms. Some of the terms used, such as "dependable motive" and "marginal man," may prove more confusing than helpful to this developing science.

An outstanding feature in the development of this text as a teaching instrument is the fact that a mimeographed edition was used and a research evaluation (to be published) made on the basis of student reactions. When will the day come for supplanting the primitive form of professional evaluation represented by this review with a research evaluation based on reactions of a sample of instructors who might use (or have used) the book as a text?

All in all, Newcomb's *Social Psychology* appears to be an able and valuable synthesis of concepts and principles drawn from general psychology, the psychology of personality, sociology, and cultural anthropology. It is a dignified and scholarly work which reflects the exacting requirements of science. Yet it is clear, vivid, and interesting throughout. Here, then, is a text well suited to win the strong endorsement of scientist, instructor, and student alike.

GLEN HEATHERS.

The Fels Research Institute.

ALLPORT, G. W. *The individual and his religion.* New York: Macmillan, 1950. Pp. viii+147. $2.50.

It may be just a coincidence, but Allport's title is reminiscent of *The Individual and His Own*, curiously enough, perhaps the most iconoclastic work produced in the nineteenth century. Max Stirner's (i.e., Kaspar Schmidt's) formidable book was more appealing to the youth of the latter half of the century than the Kinsey Report is to the present generation. Allport's title, however, has a sound reason per se, the idea behind it being the fact that religion is an individual experience and cannot be treated *en gros*. It is evident that the author is not concerned with dogmatic or organized religion but with the sentiment as such.

In his search for a definition of the religious sentiment, Allport appears to be partial to the German writers, Schleiermacher, Otto, Wobbermin, Max Weber. Yet it was in France that the psychology of religion had its beginning, and it was largely the French influence which gave impetus to William James. It is perhaps an oversight that Allport has not so much as mentioned the Franco-Jewish social philosopher, Émile Durkheim, whose large volume on religion has been a stand-by for a whole generation of sociologists, not only in the French-speaking countries. Yet we can hardly expect a complete Baedeker within the compass of restricted travel; and the author has been highly discriminating in his selection of opinion before concluding that "there are as many varieties of religious experience as there are religiously inclined mortals upon the earth"—a conclusion which will draw cavils from many quarters.

Although much of the material reviewed in the chapter on the religion of youth had been covered before in a collaborated article, the statistical account is relevant and interesting. But the results are not as encouraging as the religionist might wish to see them. In the first place, perhaps the chapter should be headed "The Religion of Social-Relations Students at Harvard and Radcliffe." Even then I am not certain whether the figures would afford us a clear-cut picture. The summation of results gained from Catholic and liberal Christian or Jewish students reminds us of the averaging of wages and salaries received by employees and high executives in General Motors.

Allport (p. 43) is at pains to gloss over the inconsistency between most of the students' believing in God of some kind and yet ruling out the supernatural. He thinks that it shows dissatisfaction with traditional ecclesiastic positions, but why are the Catholic students satisfied with their traditions? Is it not a function of (a) the kind of upbringing, (b) the degree of reflectiveness, and (c) the index of emotionality? If we had full data on these three factors, we might be able to predict the religious status of the particular student.

Has Allport asked himself what a similar questionnaire in a Moscow or Leningrad university would show, or for that matter in a Moslem institution? To my mind, it is plain that the majority of those steeped in religious rites would remain religious while a few bent on questioning everything would break away from religion, whereas those receiving a wishy-washy or lukewarm religious education would not be interested one way or another until some crisis perhaps brought out some impulse connected with self-preservation.

There is much to say on the efficacy or futility of prayer, which Allport introduces as a function in religion, but it is difficult to accept the statement that "prayer is continuous with hope as hope is continuous with prayer." Hope and fear are scarcely rectilinear but are to be envisaged in a see-saw relation. Prayer is only the *practical* resort, in either case, when the person affected is religious, not otherwise. We can hardly imagine a Tom Paine, an Ingersoll, or even a Jefferson praying when in distress.

In spite of its small compass (less than 150 pages), the book is chock-full of information and reflections set down in the best tradition of the essay. There is skillful brickwork in the rearing of the whole structure, overlaid at many points with significant utterances from the thinkers who had something to say. The elegant diction is reminiscent of the style of William James, and the whole book is richly embroidered with illustrations, which makes the little volume an example of clear exposition. The essay, too gracefully flowing to be a closely-reasoned disquisition, is a forthright examination of the whole religious situation in a psychological light. At no time is there any attempt made to by-pass the knotty problems which present themselves to the serious student.

A. A. ROBACK.

Cambridge, Massachusetts.

MUNN, N. L. *Handbook of psychological research on the rat.* New York: Houghton Mifflin, 1950. Pp. xxvi+598. $7.50.

There is a mere handful of psychologists who can give us a whole series of outstanding books. Professor Munn is one of them. This is his fourth book, and he has done it again—a monumental and magnificent job! It is scholarly, exceedingly thorough, well written, well illustrated, and very timely.

In a way this is a revision of Professor Munn's earlier book (1933), *An Introduction to Animal Psychology.* That one was about the rat, and this one is too. The first book, however, had only six hundred references, and this one has twenty-five hundred—an increase of fourfold—which is virtually all of the articles that have ever been written about the rat as an experimental subject. So Professor Munn had to make a difficult choice, whether to write a textbook based on a selected minority of research or a handbook that somehow covered them all. Fortunately for psychologists—and their students, too, I think—he chose to do the handbook. The result is an amazingly complete coverage of research on the behavior of rats.

The main text of the book is 480 pages of double-column print. It is organized into 10 chapters and 101 major subdivisions. A very attractive and well-designed table of contents lists all the divisions, major and minor, and thus makes the book much more usable. There are 180 illustrations and 15 tables. These have been selected and presented with care to show major methods and results of research. It is helpful in this kind of book that they are also listed in the front matter. Following the main text is an appendix listing more than fifty books in comparative psychology, another appendix giving many additional references to research with other vertebrate and invertebrate animals, and finally an appendix giving about twenty-five hundred references to rat behavior—a really comprehensive bibliography. The author and the publishers are to be commended especially for the very attractive way in which the book has been prepared and manufactured.

As a handbook, it will be extremely valuable for all psychologists doing research, or interested in research, with rats. Its main divisions are: (1) "Introduction" (including the care and handling of rats, and research procedures), (2) "Unlearned Behavior," (3) "General Activity," (4) "Motive, Emotions and Hoarding," (5) "Sensory Processes," (6) "The Role of Sensory Processes in Maze Behavior," (7) "The

Learning Process," (8) "Some Aspects and Conditions of Learning," (9) "Systematic Psychology," and (10) "Abnormal and Social Behavior." Under these main divisions, psychologists will find topic after topic reviewed in such a scholarly and coherent manner as they can find no other place in the psychological literature. Physiological psychologists will find details and illustrations of research otherwise found only in the original literature. Systematic and learning psychologists will find a stimulating and up-to-date account of their controversies and the experimental evidence pro and con. Even clinical and social psychologists will want to consult it for the many experiments with the rat that now bear on their problems.

It is worth emphasizing, I believe, that this book is not merely a compendium of researches. It gives coherent stories—at least as coherent as the status of research will allow. And it is to a certain degree critical, pointing out some of the shortcomings and strong points of various studies. More than that, there are some ideas in the book—ideas for further research and for the reinterpretation of past research—but the ideas are labelled clearly enough to let the reader not mistake them for facts. For all its thoroughness, then, this is not just a fact-grubbing handbook, but an integrated, critical, balanced, and interpreted account of psychological research as it is represented in the rat. It is a book, I believe, that no psychologist can afford to be without.

Even though it is a handbook, it is perhaps unfortunate that it is called one, for that may deter instructors and students from using it as a textbook. It is, indeed, an outstanding textbook. The style of writing, although it is not as elementary as in good introductory texts, is clear and interesting. Its illustrations are well chosen to represent the most significant materials.

It does, however, have some drawbacks as a textbook. One is that it is a little long for a short course in comparative psychology. Because it is so well organized, however, an instructor can select parts to leave out if he needs to cut down reading assignments. A second fault, as a textbook, is its complete coverage of the literature. References and names sometimes come along at a tremendous pace—more than the typical student wants to take. On the other hand, the writing and the continuity of the text are so good that students, with proper guidance from the instructor, can learn to skip the rapid-fire of names at the points where they occur. Thirdly, the book is exclusively about the rat, and I agree with Beach that comparative psychology *should* be evolutionary in outlook, not confined to a single species. The fact is, however, that the vast majority of our animal research is with rats and that the rat is therefore our main introduction to animal behavior. It is better to have a good book on the rat than a poor one about a lot of animals. Moreover, Munn has rendered the additional service (in an appendix) of giving a well-chosen list of references about other animals of the phyletic series, and the instructor can supplement his lectures and reading assignments from these other materials.

All in all then, although it is a handbook, it is also an excellent textbook, and I hope teachers will consider it for their courses in comparative and animal psychology. It might also serve very well in courses in advanced general psychology.

Your reviewer is so impressed with this book that he feels he can not do it justice in this brief space. Moreover, what minor faults it has do not deserve to be mentioned alongside its exceptional merits. Few, if any, authors could do what Dr. Munn has done, and none has! He has rendered an outstanding service to his colleagues, for which they certainly will be grateful. Many a student will thank him too.

C. T. MORGAN.

The Johns Hopkins University.

KATZ, DAVID. *Gestalt psychology: its nature and significance.* (2nd Ed., English trans. by Robert Tyson.) New York: Ronald Press, 1950. Pp. x+175. $3.00.

This book is, with certain exceptions, a series of 29 précis about the concepts, principles, and theoretical applications of Gestalt psychology. Katz writes clearly, succinctly, and with a mature perspective, and his translator has put his thought into equally happy English. The first half of the volume—the first 16 of his neat, well-integrated, little chapters—gives as clear, as intelligible, and as unbiased a view of Gestalt psychology as can be found today. The treatment is more extended than the histories and system-books provide, and it is less technical, less argumentative, and less interrupted by references to personal opinions of the great *Gestaltpsychologen* than are the more technical accounts which quote the literature and cite the details of research. Katz is persuasively dogmatic. He mentions particular persons and researches only casually, confining himself to what seem to him and to this reader the incontrovertible facts which have passed from the realm of personalization to the common domain of science. It is refreshing to read that kind of sure discussion in psychology.

The chapters are truly short. Nine of the 29 chapters are three pages long or less, and 24 of them are seven pages or less. These pages are, moreover, little pages—about four hundred words each. Any journalist knows that reading is made easier when subdivision is frequent and clear, and Katz himself in his discussion of mental work in this book shows that efficiency is greater when the total task is clearly separated into smaller parts. Here, however, we have no caption-writer's artificial separations, for Katz has aimed each précis at a specific goal of understanding. The reader is not left in suspense by unresolved argument nor crowded into acceptance because his doubts and reservations have become too many to remember. The author teaches him comfortably and wisely, point by point, and this reviewer profited from the instruction he received.

Katz writes from forty years' experience in the periphery of Gestalt psychology. When Wertheimer was bringing the new Gestalt psychology to birth at Frankfurt in 1910–1912, Katz was in G. E. Müller's laboratory in Göttingen working on the phenomenology of color, and he published a year before Wertheimer. So we see that the phenomenological revolt against Wundtian atomism was already in the air, an event ordained in the evolution of the *Zeitgeist*, a piece of progress for which Wertheimer's brilliance may have been a sufficient cause but hardly a necessary one. Faithfully and appreciatively Katz expounds the Wertheimer-Köhler-Koffka principles, regrets that these three have so often limited their discussion to the phenomena of vision, adds from his own researches and general wisdom the proper examples and evidence in the fields of touch, hearing, and action, and thus succeeds, by mixing appreciation with criticism, in being convincing without the appearance of special pleading.

This review is not the place to repeat the standard arguments for phenomenology and against atomism, for field theory and against production theory, for the immediate cybernetic adjustment of the organism to new conditions in accordance with physical laws and against the slow adjustment by conscious deliberation or unconscious inference. Katz has dealt well and clearly with all of these matters and more, and the reader must go to his text to hear what he says.

Katz exhibits Gestalt theory as a kind of field theory, and in this matter we become aware of another way in which Gestalt psychology was borne in upon the tide of the *Zeitgeist*. Field theory is a reaction against Newtonian mechanics and action at a distance. In physics the historical line for it is Faraday–Maxwell–Planck–Einstein. Einstein was discovering Maxwell at the turn of the century and published his special relativity theory in 1905 when he was twenty-six. Planck was at Berlin. Later Einstein, Köhler, and Wertheimer were all at Berlin. Köhler says that Planck's was the mind which influenced him most in his predoctoral days. Wertheimer knew Einstein and admired him. Gestalt psychology is a kind of relativity theory, a psychology of field patterns. Katz does not say all this, but anyone who will read his book together with P. A. Schilpp's *Albert Einstein: Philosoper-Scientist* (1949) will at once see the significance of the latter for the former.

In the second half of this book, when the discussion has passed beyond the integrations of perception and of action to the general topics of thinking, memory, volition, and emotion and to the selected applications of Gestalt theory in medical, comparative, child, and social psychology, the exposition becomes less satisfying. In part the fault here lies with psychology and not with the author. These topics have not been advanced by Gestalt principles so much as has the business of perception and motor integration. The new concepts clarify certain topics within these fields but do not reorient them, not if Katz's words about them are all there is to say. Especially in the final chapters does his selection seem arbitrary and not rounded out into the perfect précis that were his first chapters. To write the title "Medical Psychology" and then to discuss nothing but perception of form in hemianopsia, is to reduce this part of the discussion to mere incidental comment on a pertinent but adventitious topic.

Presumably the author is less at home in these later chapters, and thus the reader feels that he is being less expertly led. Let me give one example. Katz objects to associationism as dealing with separate connections and being atomistic. He gives many instances (chaps. 20–22) to show how the total constellation of the task determines learning, the formation of new mental structures. In other contexts he cites Hering and Ehrenfels as having anticipated the Gestalt psychologists in recognizing the importance of the influence of a whole field upon its parts. Why should he not, therefore, mention Ebbinghaus here instead of leaving the reader to think that nineteenth-century connectionism was thoroughly atomistic? Is there any law of Ebbinghaus' in his *Ueber das Gedächtnis* of 1885 that does not show how the total task overrules the simple law of frequency of contiguity? That it takes more repetitions to learn a longer series is a fact comparable to Hering's noting that in visual contrast the retina acts as a whole.

The other difficulty with the latter part of the book (chaps. 21–23) lies in Katz's departure from his rule of presenting only facts and principles while he omits detailed discussion of the supporting research. In these chapters he launches into description of certain experiments from his own laboratory, studies of mental work, connective inhibition, mental dazzle, and the transposition of action forms. ("Mental dazzle" is the confusion that arises when instruction or information is given too rapidly for immediate comprehension, the state from which Katz so ably protects the readers of his book.) These experiments are original, interesting, germane each to its topic, but so are the many other experiments which the author cites, some of them now become classical in Gestalt literature. A book is not a suitable place for the description of experimental techniques nor for the presentation of the raw data from which conclusions are drawn. Katz gives these results in new chapters which were added in this second edition, and in doing so he spoils the artistry of his book. If the journal sources for these researches are not easily accessible to the readers of the discussion (and they seem not to be), then the experiments could be described in appendices with only their conclusions cited in the text like the conclusions of all of the other pieces of experimental work.

The book was published first in Swedish in 1942, then in German in 1944. The second edition, radically revised all through and with four new chapters added (chaps. 13, 21–23), came out in German in 1948, and now in English with a preface dated March, 1949. The publisher has matched the author and translator in good work (though Fig. 31 is upside down), and there is a short bibliography and two indices. The text will make excellent reading for undergraduate seniors and first-year graduates in psychology in their systematic courses. Katz, at the periphery of Gestalt psychology, can see the whole of it more clearly and with less ego-involvement than those at its center, and he is thus a proper first-instructor for the novice, who can advance later to the large literature of dissent and special pleading.

Harvard University. EDWIN G. BORING.

PROTHRO, E. TERRY, AND TESKA, P. T. *Psychology—a biosocial study of behavior*. New York: Ginn & Co., 1950. Pp. ix+546. $3.75.

This book is intended as a text for the beginning course, and its announced thesis is "that a human individual is a biological organism living in a social environment, and that it is therefore necessary, if we are to understand the behavior of human beings, to study the nature of biological organisms and the nature of their social environment." Many instructors and probably most students will, however, wish that there were relatively more study of human behavior itself and less of biological structure.

Part I contains the two conventional chapters on what psychology is and on scientific method, plus a third which might better be entitled "The Concept of Self." Although this third chapter is not a philosophical discussion of the nature of the self, it is abstract and generalized to an extent that makes one wonder what has been learned from the reading of it. One point seems to be that "everyone has a highly distorted idea of what he is like," and "this distortion is almost invariably in the direction of overestimation." Consequently, "when a more objective examination is made of the individual—the interest in investigating oneself is likely to evaporate." Even were all these statements established by scientific procedures as laid down in the preceding chapter, one wonders about their effect in motivating the beginning student.

Part II consists of four chapters on the dynamic basis of behavior, including two describing the social environment as it creates and channelizes motives. The reviewer is admittedly muddled in the field of motivation, always wondering why it is better to refer to a gregarious motive than to describe common patterns of behavior. But if motives are to be discussed at any length, then the role of society in creating and channelizing them is highly pertinent. This socio-psychological treatment of motivation is a highly commendable feature of the book.

The reviewer objects to Part III, devoted to the structural basis of behavior, on two scores, but what causes him to object will cause some to commend. First, it is too lengthy and too detailed. It consumes 156 pages as compared to 158 pages devoted to behavior itself. Second, it is too unrelated to behavior, too much sheer description of structure without reference to functioning. Chapter 9 is devoted to organs of adjustment in the lower animals, a condensed course in comparative anatomy. The chapter on heredity might just as well be found in any biological text, for practically the only reference to human heredity is a short treatment of its complexity and a concluding statement that "from our study of the principles governing the transmission of traits, we can see that man's structure is largely inherited and that certain limits are set for him by the kind and quality of organism he inherits."

The five chapters of Part IV dealing with integrated behavior are "Observing," "Learning," "Efficient Learning in College," "Intelligence," and "Vocational Guidance." The logic of placing "Vocational Guidance" under a section on integrated behavior is a bit puzzling. The absence of chapters on thinking and emotions is noticeable, and some will miss the present-day treatment of personality.

The book as a whole is well written and adequately documented. The authors have carried through their intended stress on biological and social factors. Those teachers of the beginning course who emphasize these factors in their presentation will certainly want to consider adoption of this text.

C. N. REXROAD.

Stephens College.

KLINEBERG, OTTO. *Tensions affecting international understanding.* New York: Social Science Research Council, Bulletin 62, 1950. Pp. xi+227. $1.75.

Klineberg's survey of research on tensions affecting international understanding is treated in four parts: (1) personality in relation to nationality, (2) national stereotypes, (3) attitudes and their modification, and (4) influences making for aggression. The first of these topics is examined with full reference to the relevant literature. The latter three are illustrated by a sampling of representative studies. This monograph cannot serve, therefore, as a bibilography, but it is most unlikely that any of the conclusions drawn from this survey would suffer substantial modification by any research omitted from consideration.

The conclusion one must draw from this monograph is that the social sciences are ill-prepared to deal with international tensions either conceptually or pragmatically. In each of the four areas investigated by Klineberg there is a paucity of material that can be made to add up unequivocally to any general conclusion. One seeks in vain for a hard base of interrelated data or systematic theory upon which to build any programs designed to improve materially international understanding.

The following more specific conclusions seem to be warranted: (1) No empirically-based descriptions of national differences in personality have been made with sufficient reliability to produce substantial agreement among investigators. (2) Although a variety of ingenious techniques have been employed in the study of national character, there is yet to be reported a coordinated study of the same nation using these various techniques. (3) There exist no strictly comparable studies of national stereotypes from different countries. (4) The bulk of the work on modifying attitudes has been concerned with intergroup relations within a country, with little study of the modification of attitudes toward other countries. (5) At the present time it is most difficult to make specific suggestions to an international organization interested in utilizing the best methods for improving international relations. (6) Research on attitude change has been largely symptomatic and unable to deal effectively with causes. (7) Facilities for conducting international research are meagre, but those that do exist are being poorly used. Even where it is possible to conduct comparable studies in different countries, the opportunity is seldom seized. (8) Although a good deal is known about some of the influences making for aggression in the individual, the relations, if any, between individual aggressiveness and national aggression are virtually unknown.

Klineberg's monograph should serve to stress the importance of more and better organized research in this field. It should indicate what approaches have appeared to be promising and what unrewarding. Since, however, the conclusions he can draw are negative, the suspicion cannot be readily put down that the whole approach to the problems has not yet been phrased right. Proper problems for research have not yet been stated. What is now needed is an imaginative and positive proposal for a conceptually-integrated program of research that can make use of existing facilities and stimulate the creation of new ones.

University of Michigan. DORWIN CARTWRIGHT.

ZAHL, PAUL A. (Ed.) *Blindness: modern approaches to the unseen environment.* Princeton: Princeton Univ. Press, 1950. Pp. xvi+576. $7.50.

CHEVIGNY, HECTOR, & BRAVERMAN, SYDELL. *The adjustment of the blind.* New Haven: Yale Univ. Press, 1950. Pp. xvi+320. $4.00.

Vision is usually considered the sense modality with greatest significance for human adjustment to the environment. Because of this significance the loss of vision in blindness presents a wide variety of serious adjustmental problems. Further, the status of the blind in the social community, as revealed in earliest written records and as easily observed in modern culture, creates even greater difficulties for their adjustment. The two books being reviewed are important discussions of the adjustmental problems of the blind.

The volume edited by Dr. Zahl is encyclopedic in scope. While its origin is found in the activities of the Committee on Sensory Devices of the National Research Council, the final contents of the volume discuss a wide variety of problems of blindness. Thirty-one contributors, of whom ten are themselves blind, have contributed thirty-three chapters divided into seven parts: History and Welfare, Education and Psychology, Vocational Considerations, The Military Blinded, Time-Tested Boons to the Blind, Exploration of New Technological Horizons, and Ophthalmological Factors. Of specific psychological interest are chapters by Kathryn E. Maxfield on "The Preschool Blind Child"; Berthold Lowenfeld, "Psychological Foundation of Special Methods in Teaching Blind Children"; Samuel P. Hayes, "Measuring the Intelligence of the Blind"; Thomas D. Cutsforth, "Personality and Social Adjustment Among the Blind"; Edward A. Jerome and Harold Proshansky, "Factors in the Assay and Use of Guidance Devices"; and Karl S. Lashley, "Psychological Problems in the Development of Instrumental Aids for the Blind." These chapters are selected for mention because of their direct psychological content. However, the material concerned with the extent and care of the blind in the United States and other countries, problems of educational and vocational rehabilitation, methods of communication and travel aids, and the modern engineering efforts to develop electronic devices as aids in obstacle detection and in reading are of interest in considering the many problems of adjustment.

The volume by Chevigny and Braverman is a more personalized account of the adjustmental problems which face the blind. The senior author, who is himself blind, has collaborated with a clinical psychologist to explore the problems of blindness, especially as they relate to the usual social attitude toward this handicap.

In both of these volumes and especially in the second one there is evident a hope, or perhaps a demand, that the blind be given an opportunity to compete within their abilities for roles in our society. Historically the blind have been considered rather hopeless objects of charity. Chevigny in particular protests this status, and there is abundant evidence in both volumes that, given opportunity for individual activity, the blind will show achievement in many directions comparable to that of the sighted. While the problem of blindness has significant sociological and educational facets, the fundamental problems are psychological both in the areas of perception and of personality adjustment. Both books will of course be read with profit by all persons working with the blind, but they should not be neglected by psychologists working with any handicapped group, and they have much of value for clinical psychologists in general.

C. M. Louttit.

University of Illinois.

Greenblatt, M., Arnot, R., and Solomon, H. C. (Eds.) *Studies in lobotomy*. New York: Grune & Stratton, 1950. Pp. x+495. $10.00.

This book of 25 chapters summarizes the experience which was gained at the Boston Psychopathic Hospital since 1944 using translateral lobotomy on more than five hundred mental patients. The various chapters were written by different investigators and vary in their completeness and adequacy. Chapters 1, 2, 3, 4, and 11 introduce the problems and review the literature. The bibliography is very extensive. Chapters 5 and 6 are on postoperative behavior and nursing problems.

Of particular value and significance to psychologists are chapters 7, 8, 9, 10, and 12, all of which deal with various aspects of an extensive follow-up of 205 cases. Chapters 14 through 24 deal with special studies or reports from the standpoints of anatomy, cerebrospinal fluid, autonomic nervous system, bladder control, skin temperature, pain, electroencephalography, sociometric measures, and occupational therapy. Chapter 22, by C. R. Atwell, reports psychometric results on nine patients tested before and after operation with the Rorschach, sorting tests, and the Shipley test. Because of the paucity of patient material suitable for psychometric study, little was elicited by these methods.

The follow-up studies are the first of their sort which are based on a large enough group of patients so that one may have confidence in the information, which was badly needed. The report of the one- to four-year follow-up of 205 cases is by Arnot, Talbot, and Greenblatt. Social problems during the first year at home in 35 cases are reported by Broffman. Sexual behavior was investigated by Levine and Albert. The follow-up study of Arnot *et al.* made use of rating scale techniques to evaluate 30 different factors. They used the scales only when relevant before-and-after information was available. The most important change produced was the elimination of "tortured selfconcern" (fear, worry, and depression) in 87 per cent of patients studied. Of 141 patients rated on initiative the tendency was 1.6 to 1 in favor of a lower rating than had marked these patients before operation. It is pointed out that this decrease (laziness) is most marked during the first half-year, after which there is a tendency for increased initiative to appear. However, both increased age and schizoid apathy tend in many cases to fix the lost initiative at a low level.

Auditory hallucinations disappeared in 27 per cent of the patients. An improved appetite was shown by *all* patients after operation. Lack of cooperation, distrustfulness, resistiveness, and combativeness were all sharply reduced while irritability and outspokenness were increased. The vocational or work adjustment varied. In general this adjustment was not on as high a level as before the onset of illness.

Of 77 psychosurgery patients out of the hospital and living in the community, 45 were of good mental status, 23 of fair status, and 9 of poor status; of 104 patients still in the hospital, 4 were of good mental status, 27 of fair, and 73 of poor status.

"Despite many favorable signs, however, the relatives often complained, for the presence of even one unfortunate postoperative symptom in an otherwise symptom-free person could at times cripple the group activities and spoil the freedom and happiness of the entire family." The lobotomized patients in the community were not more "negative" members of society than they had been before their illness. These persons had never contributed their share to the community's activities before illness, and even among cases with best results from lobotomy many were still minus quantities. It is not the degree of assumption of financial responsibility by the patient that is disturbing so much as it is the unproductive or unoccupied attitudes.

Other traits or characteristics which have been mentioned as appearing after operation and hence probably attributable to lobotomy are alcoholic excesses, restlessness, irritability, poor appearance, overtalkativeness, deficiencies in insight, and lack of ability to do calculations. Of these only greater irritability and decreased ability to do calculations were actually increased in frequency of appearance. Other traits in the list were more often "normalized" than "abnormalized." The only one of these traits complained of by the patients was the feeling of increased irritability. In general, patients and relatives reported that preoperative moral, social, and religious attitudes continued to operate after lobotomy although many patients reported that there was a reduction in guilt feelings, modesty, and embarrassment in association with sexual activity.

Based on this survey the investigators drew up a profile or description of the "ideal" patient most likely to respond favorably to lobotomy. (Starred items below are of special importance.) Such a patient would be married, have children, be a college graduate, Protestant, hospitalized less than two years and sick less than four years after having had previous remissions from attacks of mental illness, free of organic brain damage, rapid onset,* had had electric shock therapy, had been sociable, not particularly irritable, of good family, social, and vocational adjustment. During his present illness he should have had no auditory hallucinations,* retained some insight,* had no delusions or ideas of reference, been depressed or elated but not apathetic, and have had some compulsive-obesssive* symptoms.

Drs. Greenblatt and Solomon write the summary (chap. 25) which is as fair a statement of our present-day knowledge concerning lobotomy as this reviewer has seen. This volume is recommended reading for all interested in psychosurgery and psychopathology.

CARNEY LANDIS.

Psychiatric Institute, Columbia University.

METTLER, F. A. (Ed.) (The Columbia Greystone Associates.) *Selective partial ablation of the frontal cortex.* New York: Paul B. Hoeber, 1949. Pp. xiii+517. $12.50.

This book has been hailed as a monumental work in the comparatively new psychiatric field of psychosurgery. There is no doubt that it is a fine example of cooperative research in which disciplines of psychiatry, psychology, and neurosurgery have for once let down their barriers and joined forces to consider a question which has been asked many times, namely, "What are the effects, if any, upon a patient who undergoes partial ablation of the frontal cortex?"

For practically the first time in psychosurgical research, a control group was used. However, the reviewer feels that the method of selecting the control group and the experimental group (24 cases in each group) is open to question.

The method used in selecting control and experimental groups involved a matching technique in which two lists were drawn from a master list which included age, sex, marital condition, IQ, occupation, final preoperative diagnosis, previous EST, insulin, metrazol, length of hospitalization, and previous admissions. "The judgment as to which one of the pair of matched patients was to be operated upon was made on the basis of medico-surgical and prognostic considerations, evidence of spontaneous improvement and temporary unwillingness on the part of the patient to be operated upon" (p. 21). The reviewer thinks that a better method of securing a control group is now being used in a study conducted by the staff of the Roseburg (Oregon) Veterans Administration Hospital. In this study every other or every third patient in a group of fifty will be given a pseudo-lobotomy, i.e., bone buttons are removed and replaced without completing the operation. The "two-stage operation" patient then will be given identical treatment to the "full stage" patient. He is studied for four months and at the end of this period, if indicated, the pseudo-lobotomized patient will undergo a complete lobotomy and hence will act partly as his own control.

A second question may be raised about the treatment of the control group. In the Greystone study, the control group was anaesthetized with sodium pentothal and 500 cc. of blood was taken. Theoretically this would equate two groups for blood loss but the question of knowing which group had the topectomy could hardly be disguised. This point is important since, in spite of instructions on the part of the medical staff to treat each group in the same manner, there is a good possibility that the two groups received different treatment.

Aside from this possible criticism of the experimental design, the postoperative treatment might have been changed also. This was mentioned by the experimenters; and the lack of staff for instigation of group and individual therapy was given as the reason for neglecting this aspect of treatment. The reviewer still feels that at least part of the recovery attributed to psychosurgery is due to the "total-push" type of therapy which a topectomy or a lobotomy usually brings to the patient, if for no other reason than the fact that he is receiving special attention.

Chapter 2 deals with prognostic expectation of patients under ordinary institutional care. This chapter is of great interest since it brings up the question as to spontaneous recovery in patients with psychoses of approximately the same severity as those which the patients in the control and experimental groups had. The authors state, "Bearing these facts in mind, it can be assumed with reasonable certainty that any improvement shown in the series of patients treated surgically was due to topectomy operation" (p. 29). The reviewer feels that this statement is somewhat optimistic since "total-push" therapy was not tried in the 500 cases on which these findings were based.

Section II of this work, entitled "Medical Survey," describes in detail the laboratory findings and the physical examinations of the patients in the study. As with the remaining chapters of the book, a short summary enables the reader quickly to survey the results. Of more particular interest to psychologists are the findings in this section in the fields of audition and vestibular function. The authors found that topectomy did not influence auditory function directly and was often useful in making uncooperative patients more accessible to audiographic tests.

Chapter IX, dealing with vestibular function, is of particular interest to the experimental psychologist, mainly from the instrumental point of view. The description of the Holsopple chair, used to obtain cinematic records of the nystagmus reaction, will help other experimental psychologists in dealing with this somewhat neglected aspect of perception. Holsopple has devoted many years to this sensory process and at last has succeeded in overcoming the technical difficulties always present in previous rotational studies, namely that of controlled acceleration. In general, the authors found that topectomy resulted in lengthening of nystagmus time and that there was a positive relationship ($r = +.75$) between the amount of tissue removed and the objective scores. They observe, however, that more extensive testing should be done before this evidence can be accepted as conclusive.

Chapter X, dealing with the visual apparatus, reveals that in the present sample there is no effect of partial ablation of the frontal cortex on visual acuity, acuity of peripheral vision, and color vision. Interestingly, there is some change noted in the autokinetic phenomenon; it appeared to be absent three weeks postoperatively and then recovered three months postoperatively. This recovery is not complete; the patterns are "simpler" after the operation.

Section III is of most particular interest to clinical psychologists. As the author of this section states, "The primary question for which we sought an answer was: 'Does excision of frontal lobe tissue influence behavior and mental life?'" They find, as might be expected, that psychologic and psychometric changes do take place; there was no real loss of learning, memory, or intellectual function; there was a statistically significant gain in recall and perceptual recognition. Loss of ability to do abstract thinking apparently was not demonstrated. Apparently psychologic changes depended to some extent on the areas involved. Areas 9, 10, and 46 are associated with social recovery. Ablation of Area 11 showed little or no change in behavior. Area 44 removal was not related to changes in verbal or speech function. In reviewing the areas ablated in the Greystone studies, it is interesting to note that the anterior cingulate gyrus (Area 24) was not reached except in one case; A. A. Ward and K. E. Livingston, two neurosurgeons, feel that this area may well be related to behavioral changes. Evidence to date indicates this to be true in monkeys.

The remaining chapters in Section III deal with several new psychologic techniques used for the first time with psychosurgical patients. To measure changes in affectivity, one of the widely-discussed effects of such operations, M. Garrison devised two questionnaires which attempt to measure anxiety and complaints. These questionnaires (30 questions each) were answered by the patient, and responses were weighted. Results were interpreted cautiously and, as the investigators state, represent tendencies only. They found, for example, that anxiety as measured by their questionnaires tended to decrease in patients with Areas 10 and 46 removed. This decrease was related to social improvements. Complaints as measured by the second questionnaire also decreased.

In addition to these inventories, the Mirror Drawing task was used as a measure of affectivity. Using blocking (defined as interruption of tracing for a period of time) and frustration (measured subjectively on a five-point behavior scale), they found a tendency for blocking to decrease when Area 9 was excised with patients of average intelligence. Frustration did not seem to be related to any of the criteria considered.

Radical surgery of the frontal lobes was not reflected in changes in Rorschach protocols; no definite patterns of changes could be observed in the operative group.

Section IV deals with the results of the experiment in terms of psychiatric, neurologic, and pathologic changes as seen before and after topectomy and lobotomy. The psychiatric findings in general are not too encouraging. The experimenters in this section are cautious in making any interpretation, because of the small sample. However, they do state definitely that there was no relationship between weight of brain removed and improvement as measured by behavior scales, and that therapeutic response did not seem to be related to type of onset of illness or precipitating causes. Favorable response to previous electro-shock therapy (EST) tended to be predictive of good results of the operation. Lastly they found that clinical improvement was best in younger patients whose psychosis was of short duration. This point is interesting since it shows more than ever the need of careful control groups. Accounting for spontaneous remissions in younger patients with short psychotic histories has been the bane of investigators in the field of psychiatric research for many years.

Fully aware of the fallacies of nosological classification, the authors make a few tentative statements about the type of illness which responds to psychosurgery. These are of interest, particularly if one considers that lobotomy has been recommended in anxiety neurosis. Although no psychoneurotic patients were included in this sample, the authors of this section state, "One should be hesitant to recommend operation for only but the most severe neurotics and then only if intensive and skillful psychotherapy has proved valueless" (p. 428). They feel that only schizophrenics with affect somewhat appropriate to ideation should be used as subjects. This obviously includes the better integrated schizophrenic with whom, naturally, results would be more striking. They feel that a manic-depressive in the depressed phase is a good candidate. They do not have evidence on psychopaths but feel that tension in this type of personality is not of the same kind as that relieved in psychotic patients. They do not feel that surgery can replace EST as a treatment of choice. Lastly, the improvement resulting from operative procedure is basically due to a change in affective response, and no patient was actually worse after the operation. Some cases showed temporary improvement only to have a relapse.

We see then that the effects of the removal of the frontal lobe tissue are not definitely established, although the Greystone study has added immeasurably to our knowledge of this problem and has uncovered new problems. The reviewer believes that improvement of the control group by use of pseudo-lobotomy and "total-push" therapy, including individual and group therapy, will yield additional knowledge concerning the effects of lobotomy. The Greystone study does show that selection of patients for lobotomy is still a difficult problem and that it should not be taken as lightly as it has been in the past. "The psychiatrists' therapeutic armamentarium," to quote Heath, still includes intensive therapy.

FRAMPTON B. PRICE.

Veterans Administration Hospital, Roseburg, Oregon.

BOOKS AND MATERIALS RECEIVED

ADINARAYAN, S. P. *The human mind.* London: Hutchinson Univ. Library, 1950. (New York: Longmans, Green.) Pp. 195. $1.60 text; $2.00 trade.

BODER, DAVID P. *I did not interview the dead.* Urbana, Ill.: Univ. of Illinois Press, 1949. Pp. xix+220. $3.50.

BURACK, BENJAMIN. *The nature and efficacy of methods of attack on reasoning problems. Psychological Monographs* No. 313 (Vol. 64, No. 7). Washington, D. C.: American Psychological Assn., 1950. Pp. v+26. $1.00.

CAMPBELL, PERCY A. *Our natural universe—including man.* Philadelphia: College Offset Press, 1950. Pp. iv+75. $2.00.

CANTRIL, HADLEY (Ed.) *Public opinion 1935–1946.* Princeton, N. J.: Princeton Univ. Press, 1951. Pp. lix+1191. $25.00.

CANTRIL, HADLEY. *The "why" of man's experience.* New York: Macmillan, 1950. Pp. xiii+198. $2.75.

CATTELL, RAYMOND B. *An introduction to personality study.* New York: Longmans, Green, 1950. Pp. vii+235. $1.60 text; $2.00 trade.

DOCKERAY, FLOYD C., AND LANE, GORHAM G. *Psychology.* (2nd Ed.) New York: Prentice-Hall, 1950. Pp. viii+564. $3.75.

DOLLARD, JOHN, AND MILLER, NEAL E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950. Pp. xiii+488. $5.00.

ERON, LEONARD D. *A normative study of the Thematic Apperception Test. Psychological Monographs* No. 315 (Vol. 64, No. 9). Washington, D. C.: American Psychological Assn., 1950. Pp. v+48. $1.50.

FERENCZI, SANDOR. *Sex in psychoanalysis.* New York: Basic Books, 1950. Pp. 338. $3.50.

FISCHER, SIEGFRIED. *Principles of general psychopathology.* New York: Philosophical Library, 1950. Pp. xxi+327. $4.75.

FRISCH, KARL VON. *Bees; their vision, chemical senses, and language.* Ithaca, N. Y.: Cornell Univ. Press, 1950. Pp. xiii+119. $3.00.

GIBSON, JAMES J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950. Pp. xii+242. $4.00.

GOULDNER, ALVIN W. (Ed.) *Studies in leadership. Leadership and democratic action.* New York: Harper, 1950. Pp. xvi+736. $5.00.

GRECO, MARSHALL C. *Group life: The nature and treatment of its specific conflicts.* New York: Philosophical Library, 1950. Pp. xvi+357. $4.75.

GRIMES, JOHN MAURICE. *When minds go wrong. A simple story of the mentally ill—past, present and future.* Chicago: Author, 1949. Pp. 237. $5.00.

GULLIKSEN, HAROLD. *Theory of mental tests.* New York: Wiley, 1950. Pp. xix+486. $6.00.

Kaufmann, Walter A. *Nietzsche: philosopher, psychologist, antichrist.* Princeton, N. J.: Princeton Univ. Press, 1950. Pp. xi+409. $6.00.

Korner, Ija N. *Some aspects of the problem of repression: Repressive forgetting.* New York: Bureau of Publications, Teachers Coll., Columbia Univ., 1950. Pp. 60. $2.10.

Lord, Edith. *Experimentally induced variations in Rorschach performance. Psychological Monographs* No. 316 (Vol. 64, No. 10). Washington, D. C.: American Psychological Assn., 1950. Pp. 34. $1.00.

Meehl, Paul E. *Using the Minnesota Multiphasic Personality Inventory in counseling; a summary of selected new research results.* St. Paul, Minn.: Advisement & Guidance Section, VRED, VA Center, Fort Snelling, 1950. Pp. 42. (Mimeographed.)

Meyer, Max F. *How we hear. How tones make music.* Boston: Charles T. Branford, 1950. Pp. 117. $2.50.

Morland, Nigel. *An outline of scientific criminology.* New York: Philosophical Library, 1950. Pp. 287. $4.75.

Mowrer, O. Hobart. *Learning theory and personality dynamics.* New York: Ronald, 1950. Pp. xviii+776. $7.50.

Nuttin, Joseph. *Psychanalyse et conception spiritualiste de l'homme.* Louvain, Belgium: Editions de l'Institut Supérieur de Philosophie, 1950. Pp. 436. $2.80. 140 fr. b.

Persons, Stow. (Ed.) *Evolutionary thought in America.* New Haven: Yale Univ. Press, 1950. Pp. x+462. $5.00.

Pfeiffer, John. *Genetics—the science of heredity.* Public Affairs Pamphlet No. 165. New York: Public Affairs Committee, 1950. Pp. 32. $.20.

Piaget, Jean. *The psychology of intelligence.* New York: Harcourt, Brace, 1950. Pp. viii+182. $3.50.

Rashevsky, Nicolas. *Mathematical biology of social behavior.* Chicago: Univ. of Chicago Press, 1951. Pp. xii+256. $5.00.

Remmers, H. H., Drucker, A. J., and Christensen, H. T. *Courtship conduct as viewed by high school youth.* (Report of Poll No. 27 of The Purdue Opinion Panel, Vol. 10, No. 2, December, 1950.) Lafayette, Ind.: Division of Educational Reference, Purdue Univ., 1950. Pp. 37. (Mimeographed.)

Riezler, Kurt. *Man: Mutable and immutable.* Chicago: Henry Regnery, 1950. Pp. x+365. $5.00.

Rogers, Carl R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951. Pp. xii+560. $4.00.

Rosenblueth, Arturo. *The transmission of nerve impulses at neuroeffector junctions and peripheral synapses.* New York: Wiley, 1950. Pp. xiv+325. $6.00.

Scheinfeld, Amram. *The new you and heredity.* Philadelphia: J. B. Lippincott, 1950. Pp. xxii+616. $5.00.

Selover, Margaret, Townsend, Agatha, Jacobs, Robert, and Traxler, Arthur E. *Introduction to testing and the use of test results.* Educational Records Bureau, August, 1950. Pp. vii+107.

Simpson, Elizabeth A. *SRA Better Reading Book 2.* Chicago: Science Research Associates, 1950. Pp. 88. $1.55.

Singer, Martin. *The validity of a multiple-choice projective test in psychological screening. Psychological Monographs* No. 314 (Vol. 64, No. 8). Washington, D. C.: American Psychological Assn., 1950. Pp. v+40. $1.00.

Stolzenberg, Jacob. *Psychosomatics and suggestion therapy in dentistry.* New York: Philosophical Library, 1950. Pp. xi+152. $3.75.

Stone, Calvin P. (Ed.) *Annual review of psychology.* (Vol. II.) Stanford, Calif.: Annual Reviews, 1951. Pp. ix+389. $6.00.

Templin, Mildred C. *The development of reasoning in children with normal and defective hearing.* (Institute of Child Welfare Monograph No. 24.) Minneapolis, Minn.: Univ. of Minnesota Press, 1950. Pp. xii+143. $3.00.

Thompson, Laura. *Culture in crisis: A study of the Hopi Indians.* New York: Harper, 1950. Pp. xxiv+221. $4.00.

Thurstone, L. L. *Thurstone temperament schedule* (with manual). Chicago: Science Research Associates, 1950. $.75.

Tupes, Ernest C. *An evaluation of personality-trait ratings obtained by unstructured assessment interviews. Psychological Monographs* No. 317 (Vol. 64, No. 11). Washington, D. C.: American Psychological Assn., 1950. Pp. vi+24. $1.00.

Wright, Herbert F., and Barker, Roger G. *Methods in psychological ecology; a progress report.* Lawrence, Kan.: Field Study of Children's Behavior, Univ. of Kansas, 1950. Pp. ix+273. $2.50.

Bulletin analytique de documentation politique, économique et sociale contemporaine. Fondation Nationale des Sciences Politiques. Paris, France: Presses Universitaires de France, 1950. Pp. 48. 1000 fr.

The exceptional child in infancy and early childhood. Proceedings of the Annual Spring Conference on Education and the Exceptional Child. Langhorne, Pa.: Child Research Clinic, The Woods Schools, 1950. Pp. 48.

A survey report on human factors in undersea warfare. Panel on Psychology and Physiology. Washington, D. C.: Committee on Undersea Warfare, National Research Council, 1949. Pp. x+451. $2.25.

Research relating to children: An inventory of studies in progress. (July 1, 1949–March 31, 1950.) Clearinghouse for Research in Child Life. Washington, D. C.: Federal Security Agency, Social Security Administration, Children's Bureau, 1950. Pp. 374 A–316 F+xv.

Vol. 48, No. 3 May, 1951

Psychological Bulletin

SALIVARY pH AS A PSYCHOPHYSIOLOGICAL VARIABLE[1]

J. C. GILCHRIST[2] AND ERNEST FURCHTGOTT[3]

University of California, Los Angeles

There are many suggestions in the literature of the biological sciences that the pH of saliva[4] may be a valuable tool for the diagnosis and prediction of behavior. These suggestions arise from two major considerations. First, the salivary glands are dually innervated by the sympathetic and the parasympathetic nervous systems and, consequently, may furnish indices of autonomic balance. Secondly, it is well known that extreme shifts in the pH of blood bring about behavioral changes ranging from tetany in alkalinosis to coma in acidosis. The blood, however, is well buffered, and minor shifts in the organism's acid-base balance do not appear in blood pH determination. Since saliva is less well buffered, it seems possible that its pH may serve as an index of the acid-base balance of the fluid matrix. Wenger's (58) recent study of Air Forces personnel supports at least the first of these two possibilities. That study showed salivary pH to be significantly related to other autonomic nervous system functions and demonstrated that it can be used to estimate autonomic balance. Such findings indicate that a review of the literature pertaining to salivary pH may be valuable.

Starr (49), in his paper in 1922, presented the results of earlier investigators dating back to Donné's observations in 1833. The present review attempts to cover the work done in the field since 1922. Particular attention is devoted to recent reports which, since the advent of electrometric methods, provide the most reliable data.

---

[1] The authors wish to express their appreciation to Dr. M. A. Wenger for suggesting this review and assisting them in its preparation.

[2] Now at the University of Wisconsin.

[3] Now at the University of Tennessee.

[4] pH refers to the relative concentrations of H^+ and OH^- ions within a substance and thereby its actual acidity or alkalinity. The pH value is the negative logarithm (the sign is dropped for convenience) of the concentration of the hydrogen ion in gram atoms per liter. A pH of less than 7 denotes acidity, whereas a pH of greater than 7 denotes alkalinity.

Psy. Research
COLLEGE

METHODS OF COLLECTION AND MEASUREMENT

It should be pointed out at once that a majority of reports dealing with determinations of salivary pH have made use of data obtained from mixed whole saliva. A few investigators, however, have collected the secretions of individual glands. In most cases they worked with the parotids and employed a Lashley disk (**34**) attached by negative pressure over the orifice of the gland. There are several modifications of this disk, such as an added third baffle, used by Moore and Marcuse (**38**), but it is doubted that the use of this or other variations perceptibly increases the validity of the method. In a recent study, Schmidt-Nielson (**45**) used two rubber cups which were attached to the membrane around the papillae of the parotid ducts, while the submaxillary flow was sucked directly from the floor of the mouth just under the tongue. When mixed whole saliva is collected, it is usually done by use of negative pressure devices or by having the subject expectorate.

Saliva from individual or all glands is usually collected under oil, toluene, or some other inert substance to prevent the loss of CO_2, since it was shown by Starr (**50**) that this loss causes a rise in salivary pH. Schmidt-Nielsen (**45**) notes that the loss of CO_2 from saliva is the greatest in the first 90 seconds and that toluene or paraffin oil are not very effective in preventing this loss inasmuch as they have a CO_2 diffusion rate only four times slower than water. Wenger, Furchtgott, and Voas (**59**) measured electrometrically the salivary pH of 12 subjects, with the saliva collected first under toluene and later measured directly in the mouth. For the direct measurements the mean and standard deviation were 6.45 and 0.36, respectively, and under toluene they were 6.50 and 0.40, respectively. The difference involved is not statistically significant. Schmidt-Nielsen used a syringe pipette into which saliva was sucked directly from a rubber cup and the dead space in the syringe was filled with mercury to prevent CO_2 loss. Most workers have not controlled the initial stages of their collection so carefully, and it will be seen later that they have obtained pH values higher than those obtained by Schmidt-Nielsen. Certain varying interpretations of results become possible if these factors are kept in mind, but lack of rigorous initial control need not invalidate the data from a single study as long as the worker maintains a constant technique. In fact, if it be assumed that the initial CO_2 loss is similar for all samples, it has been shown (**59**) that if pH determinations are made within five minutes after collection, it is unnecessary to use any protecting device for CO_2, unless an accuracy of greater than 0.01 pH units is desired. Gaskill (**22**) even reports that saliva stored in a small Pyrex bottle fitted with an airtight cap remains unchanged in pH for 48 hours.

As Starr (**51**) and others have pointed out, neither titration nor colorimetric indicators provide entirely satisfactory salivary pH determinations. Titration must be questioned because the concentration of hydrogen ions in a given volume of liquid, rather than the quantity of alkali replaced in the course of titration, determines the significant biological properties of most fluids. The efficiency of the liquid colorimetric indicators is impaired due to the wide variation in the turgidity of saliva. This, of course, makes for varying degrees of chromatic saturation in the indicator, since these methods require the use of a color comparator with the same light source used for illuminating both the standards and the material to be tested.

Glaubiger and Falk (**23**) concluded from a series of investigations using electrometric and colorimetric techniques that the latter may be used if they are checked regularly with standard solutions and if determinations are needed only within 0.3 pH. Paper colorimetric indicators, in the main, are designed to measure only within the nearest 0.5 pH unit and have only a short range. The Wulff paper indicator strip is designed to be used for turbid, viscous materials with an accuracy of 0.1 pH unit, although by the use of standard buffer solutions it is claimed that the accuracy can be extended to 0.05 pH unit. No data have been found which compare this technique with other colorimetric methods.

Most investigators regard electrometric techniques as the only ones providing accurate pH determinations, and the glass electrode is the most frequently used today. Soyenkoff and Hinch (**48**) claim that the difference in using a quinhydrone electrode is less than 0.02 pH, but Wessinger (**56**) notes that the potential drift is greater with increased deviation of the oxidation-reduction potential from zero. He describes a vacuum tube potentiometer which eliminates this drift.

Microtechniques are now being used by some workers, and studies in situ appear possible. Schmidt-Nielsen (**45**) recently published a method for microdetermination which produces fairly reliable results with less than 0.10 ml. of saliva required for determination. A Billman-type (**7**) quinhydrone capillary electrode is employed, filled with about 0.05 ml. of a standard buffer solution. Claff and Swenson (**15**) describe another technique for the determination of samples of biological fluids as small as 0.005 ml. Nims's (**40**) glass electrode and apparatus used to measure the pH of the brain surface in vivo could be adopted to determine the pH of salivary glands in situ. Stephan (54) has employed an antimony electrode to measure the pH of the tooth surface, which might also be modified to measure salivary pH. Recent work (59) indicates that direct oral measurements are possible with small commercial glass electrodes.

pH VALUES FOR HUMAN SALIVA

Until recently, most authors have considered saliva as having an alkaline reaction. (See Zoethout and Tuttle [62, p. 343] for a typical statement of this position.) Table I shows the results obtained in ten of

TABLE I

REPORTED VALUES OF SALIVARY pH

Investigator	*Collection Method*	*Measurement Method*	*Samples and Subjects*	*pH Values Reported*
Brawley (12)	Expectoration	Colorimetric (liquid)	1 sample from each of 1826 males and 1578 females, aged 3 weeks–101 years	$M = 6.75$ (Males $= 6.76$; Females $= 6.73$) $Mdn = 6.84$
Carlson & McKinstry (14)	Expectoration	Colorimetric (liquid)	1 sample from each of 127 male college students	$M = 6.85$ Range $= 6.66$ to 7.02
Eisenbrandt (19)	Expectoration	Electrometric	276 samples from 5 subjects at 9, 11, 1, 3, and 5 o'clock during the day	$M = 6.72$, $\sigma = .242$
Eisenbrandt (20)	Expectoration	Electrometric	1552 samples from 7 subjects at hourly intervals during the day for 1 week of each month for 1 year	$M = 6.64$, $\sigma_M = .006$
Gans (21)	Expectoration	Colorimetric (liquid)	1 sample from each of 90 subjects	Range $= 6.0$ to 7.0
Reindel (42)		Colorimetric	50 cases with "normal" gastric acidity	39 of 50 cases ranged between 6.9 and 7.1
Rich (43)	Expectoration	Electrometric	3 samples from each of 57 subjects 2 hours after meals	$M = 6.82$ Range $= 5.88$ to 7.38, with 31% of observations > 7.0
Starr (50)	Expectoration	Colorimetric (liquid)	610 samples from 228 subjects	$M = 6.6$ $Mdn = 6.6$ Range $= 5.95$ to 7.25, with 86.6% of the samples between 6.55 and 7.00
Wenger (58)	Drawn from oral cavity under negative pressure	Colorimetric (paper)	1 sample from each of 488 aviation students; 225 operational fatigue cases; 98 hospitalized psychoneurotic patients; 166 army returnees without operational fatigue	$M = 7.2$, $\sigma = 0.4$ $M = 7.1$, $\sigma = 0.5$ $M = 6.9$, $\sigma = 0.7$ $M = 6.9$, $\sigma = 0.6$
Wenger, Furchtgott, & Voas (59)	Drawn from oral cavity under negative pressure	Electrometric	1 sample from each of 69 college students (both sexes)	$M = 6.45$, $\sigma = 0.41$

the most complete studies of mixed unstimulated[5] saliva in normal human subjects. Although several of the studies are incompletely described statistically, and only eight of the ten report even so much as the mean, seven of these eight studies report fairly consistent results. The obtained means were 6.75, 6.85, 6.72, 6.64, 6.82, 6.78, and 6.45. Three of these studies used colorimetric techniques, while four sets of data were obtained electrometrically. Of the remaining experiments only one reports values consistently on the alkaline side. This group of studies indicates that the pH of mixed unstimulated human saliva is slightly acid, and places its actual mean value at some point between 6.45 and 6.85. The higher values reported by Wenger (58) remain to be explained.

Little evidence is available on the comparative contributions of the different glands. The Schmidt-Nielsen study (46) provides the only available comparison of parotid and submaxillary pH measured in the same subjects. The average parotid pH for seven subjects, tested four times daily, and measured electrometrically, was 5.81, and that for submaxillary saliva was 6.39. Winsor and Korchin (60), also using electrical methods, report 5.83 as the mean parotid pH for several hundred subjects. As we have seen, these are considerably lower values for mixed saliva than those of other investigators when similar measuring techniques are employed. It would seem that the various salivary glands make differing contributions to the whole, a conclusion which is not surprising since the parotids are composed largely of serous-type cells, while the submaxillary and sublingual glands contain mucous-type cells. It remains for future research to determine the differential participation of the different glands on the results obtained for mixed saliva.

SEASONAL VARIATIONS

In addition to the problem of measurement itself, there are questions concerning the effects of experimental conditions on salivary pH. One of the questions is that of seasonal variations. Eisenbrandt (20), measuring 1552 specimens from seven subjects tested at bihourly intervals for one week of each month for one year, found by an analysis of variance that the pH of the group was significantly variable from month to month at each bihourly period.

While seasonal variations were observed in the above study, no attempt was made to correlate them with atmospheric conditions. Indirectly, a study by Berg *et al.* (3) may be compared. They found that the pH of venous blood measured in a longitudinal study of six dogs and six human beings for one year correlated with a significance above the 1 per cent level with barometric pressure and air temperature. The correlations were:

---

[5] "Unstimulated" means here that the salivary flow was not elicited by the injection of stimulants intravenously or by placing a stimulating material in the oral cavity.

	Barometric Pressure	*Air Temperature*
Human venous blood.....	.77	.47
Canine venous blood......	.69	.45

The relationship between air temperature and pH was marked only at higher temperatures. Bearing more directly upon the problem is the preliminary study by Wenger (58) on 180 Air Forces cadets over a four-month period. He obtained a correlation of +.21 between salivary pH and air temperature, and +.06 between salivary pH and barometric pressure when the relationship between barometric pressure and air temperature was partialed out.

Insofar as there is a positive relationship between salivary output and pH, the effect of seasonal changes on the former is pertinent to the review. An earlier study by Wenger (57) on 51 children and again on 23 subjects (57) reported greater salivary output in the summer than in the winter; however, the differences were not statistically significant (*t* values reported were 1.40 and 0.97, respectively). Wainwright and Becks's study (55) also furnishes evidence on this point. They investigated rate of flow of saliva in 50 subjects over a period of three years. There were no seasonal changes found. From these studies it can be concluded that there is probably only insignificant seasonal variation in salivary pH except at higher temperatures, where changes can be expected due to the physicochemical relationships between temperature and pH.

Diurnal Variations

Henderson and Millet (27) state that saliva is most acid in the morning upon arising, begins to approach neutrality immediately, rises to alkalinity after breakfast, drops to acidity afterwards, rises again after ingestion of the noon meal, following which there is another drop. The determinations were done colorimetrically using the expectorated saliva of 13 subjects who were tested during the course of a whole day, and numerous others who were tested for half-days during a period of 15 days. McClelland (36) previously found somewhat the same type of diurnal variation in four subjects tested on four days, although his absolute values differ from those of Henderson and Millet (27). Measuring expectorated saliva colorimetrically, he found the lowest pH to occur after arising, followed by a gradual increase reaching its peak just after lunch, which was followed by a decline of the pH values. Smith

(**47**), who employed a quinhydrone electrode on an unspecified number of subjects, obtained data similar to those of Henderson and Millet (**27**). The lowest pH was recorded in the morning; this was followed by a rise reaching its maximum about four hours after breakfast and then the pH dropped again.[6] Eisenbrandt (**19**) examined 276 specimens of five subjects electrometrically and found that the lowest pH occurred in the morning, followed by an increased pH at 11 A.M. and a peak at 5 P.M. Grossman and Brickman (**25**) studied three groups which were differentiated on the basis of diurnal and nocturnal activity. Group 1, active at night, had a diurnal pH of 6.7 and a nocturnal pH of 6.7. Group 2, which had eight hours of sleep at night, had a diurnal pH of 6.7 and a nocturnal pH of 6.1. Group 3, consisting of ambulant and bed patients, had a diurnal pH of 6.3 and a nocturnal pH of 6.1. The saliva consisted of mixed expectorated specimens, and the pH was determined with a quinhydrone electrode. In contrast to these studies, Schmidt-Nielsen (**46**) using his technique which has been described previously, examined the parotid saliva of five subjects four times daily and the saliva of two subjects seven times daily at regular intervals during the 24-hour period and found no diurnal or nocturnal variations in the pH of the saliva. However, he did find fluctuations in the pH of submaxillary saliva. Wenger (**58**), in his study of 180 Air Forces cadets, obtained an *r* of +.19 between time of day and salivary pH. In a more recent study, Wenger, Furchtgott, and Voas (**59**) found that with 39 subjects whose salivary pH was measured at 10 A.M. and again at 2 P.M., the difference between the means was insignificant (C.R. = .5) and that the two measurements correlated with an *r* of +.69. No attempt was made to control the subjects' usual eating habits, but subjects were instructed to refrain from eating for two hours prior to measurement, and presumably most of the measurements were obtained after the ingestion of the morning and the noon meals, respectively.

Although these results are not entirely consistent, it would seem that when diurnal variations in salivary pH are found, the values are not specifically related to the time of day but, rather, to physical activity and to the ingestion of food, both of which may increase pH temporarily.

DIET

Diet apparently has no effect upon salivary pH. Smith (**47**) fed an unspecified number of subjects a high protein diet, a diet high in $NaHCO_3$ and sodium citrate, and a high carbohydrate diet. The salivary pH of the subjects was independent of the diet fed, in contrast to the urinary pH, which was affected by the diets. Clark *et al.* (16) found that the acid-base balance of the diet bore little or no relationship to the quantity of acid or base found in the blood or in unstimulated saliva. Schmidt-Nielsen (46), measuring salivary pH soon after the ingestion of a meal, found that it was independent of the character of the meal.

---

[6] Lashley (**34**) found that the presence of food in the stomach increases salivary flow. Coupled with the relationship between rate of flow and pH (see pp. 201–202) this finding would account for the peaks in the above studies.

CHEMICAL STIMULANTS IN THE ORAL CAVITY

The results of placing various chemical substances in the oral cavity are consistent in their effects upon salivary pH. The inverse relationship of salivary pH to the pH of the stimulating substance has been demonstrated in animal and human subjects by Popow and Kudrjawzew (41), Bienka and Sczepansky (5), Bienka (6), Krassusky *et al.* (33), and Winsor and Korchin (60). These latter authors, however, state that the salivary pH remained directly proportional to the rate of flow of saliva, and thus the pH of the stimulating substance may have only an indirect effect on salivary pH by controlling salivary flow which, as will be described later, bears a direct relationship to salivary pH.

JAW MOVEMENTS

Many investigators have noted that jaw movements increase salivary pH. McClelland (36), who used colorimetric measures of expectorated saliva, found an increase of pH after movements of the jaw. Sneezing, coughing, and talking produce a similar rise. Schmidt-Nielsen (46) recorded an increase from 5.62 to 6.68 pH units after coughing. Carlson and McKinstry (14) also found an increased pH after jaw movements, while Winsor and Korchin (60), working with parotid glands, found that the jaw movement, regardless of the nature of the chewed object, produced an increased rate of flow which was almost directly proportional to pH.

Wenger, Furchtgott, and Voas (59) measured electrometrically the pH of saliva in the oral cavity for 69 subjects and compared the mean value obtained by direct measurement with that obtained by a vacuum suction collection method in which chewing movements were involved. The mean of the former was 6.36, and that of the latter was 6.45. Although the difference is not statistically significant, it is in the expected direction.

RATE OF SALIVARY FLOW

As was indicated above Winsor and Korchin (60), working with human subjects aged 13–23, found a direct relationship between rate of flow and salivary pH. Determining the pH electrometrically, they found an increased rate of secretion of the parotid gland is paralleled by an increased salivary pH. Conversely, a decreased rate of parotid secretion also produces a more acid saliva. Wenger (58) also obtained a positive correlation (r of +.44) between salivary pH and output in his study of 488 aviation students, and for 201 *Operational Fatigue* patients, an r of +.30 was obtained. All pH determinations in this study were by means of nitrazene paper indicators.

These results agree with the observations that, in acidosis (4) and fatigue (1), when the blood lactate is high and there is an interchange between this substance and the gland bicarbonate, the rate of secretion is low and the pH is also decreased.

The above results are at variance, however, with those obtained on subhuman subjects. Brassfield (8) found that the secretory rate in dogs is inversely proportional to pH. Increasing the cellular acidity of the gland inhibits the destruction of acetylcholine, resulting in an increased secretion. The effect is similar to observations in other cholinergic systems, according to Brassfield, who further states that "variations in base with variations in rate of secretion, the type of metabolism within the gland, are all important factors in determining the reactions of saliva" (p. 48). The work of Bienka (5), Bienka and Sczepansky (6), and Bronk and Gesell (13) agrees with that of Brassfield—i.e., that pH is inversely proportional to rate of flow. Kesztyüs and Martin (30) also report that increased secretion is accompanied by a decreased content of Na and therefore of total base, which means that the pH will be lowered. The above studies were conducted primarily on the submaxillary glands of dogs.

Two possible reasons present themselves for the discrepancies between the human studies and those on other animals. The first is that the studies of nonhuman animals elicited saliva by artificial means (e.g., drugs, electrical current, etc.) which may produce different results. In fact, Baxter (2) has demonstrated that the composition of saliva after pilocarpine stimulation is different from that obtained by the use of an electrical current.

Furthermore, the studies on human subjects employed either "mixed saliva" or secretions of the parotids only, while the nonhuman data were obtained from the submaxillary glands. The differences, therefore, may be due to the differential activity of these glands upon stimulation. Stavraky (53), for example, has shown that the two glands react differently to adrenaline following the previous administration of pilocarpine; and it is even more appropriate to note the results of Holzlöhner and Airapetianz (38), who found that sympathetic stimulation of the submaxillary glands in dogs produced an increase in secretion. This is in contrast to the parotids, where sympathetic stimulation results in an inhibition of salivary flow. The above findings, taken in conjunction with the work of Kesztyüs and Martin (31) (who report a lower pH for sympathetically than for parasympathetically stimulated submaxillary saliva in dogs), give a plausible explanation for the apparent contradictions between the human and other animal studies as regards the relationship between rate of flow and pH. The direct relationship between rate of flow and pH observed in human subjects is probably due to an increased pH with parasympathetic stimulation of the parotids, which also produces an increased rate of flow. The inverse relationship between rate of flow and pH in dogs is due to sympathetic stimulation of the submaxillary glands, which raises the rate of flow while at the same time it lowers the pH. From the available data it would appear, therefore, that the relationship between rate of flow and salivary pH is an indirect one, expressing mainly the increased pH resulting from parasympathetic stimulation, which in the case of the parotids or mixed saliva is also associated with increased rate of flow.

pH OF THE BLOOD AND OTHER BODY FLUIDS

Hugh and Marenczi (29), Gregersen and Ingalls (24), and DeBeer and Wilson (18) found a rise in salivary pH in animals following a slow or moderate intravenous injection of hydrochloric acid. However, according to Benetate and Oprisiu (4), when the concentration of acid is increased, or in acidosis, or after exercise, the pH of saliva follows that of blood, i.e., it drops. Leung (35) also reported that the pH of human saliva decreases under a lowered oxygen tension, and similar results were previously reported by Starr (50).

Brassfield and Hong (11) found no change in either flow or pH after intravenous injections of lactic and hydrochloric acids, while sodium carbonate and sodium bicarbonate decreased the flow and increased the pH of dogs' submaxillary saliva, and ammonium hydroxide and ammonium carbonate increased the flow and decreased pH, according to Krassusky *et al.* (33) and Brassfield *et al.* (10). The latter authors relate the apparently contradictory results obtained with the ammonium compounds to their anti-cholinesterase activity. Previously, Brassfield and Hands (9) reported the same conclusions. They also noted that procedures which caused a greater pH increase in blood than in saliva also decreased the salivary flow, while in those cases in which there was a sharper rise in the salivary than in the blood pH, the rate of salivary flow was increased. Overventilation, for example, produced an increase in both blood and salivary pH, with a lesser increase in the former and an accompanying increased salivary flow. Again, in 1945, Brassfield (8) described some experiments in which he stimulated the submaxillary glands of dogs continuously with pilocarpine and confirmed his earlier findings. He says of this work: "The lowering of the CO_2 tension of the blood increases the CO_2 gradient between the gland cell and the blood, thereby providing a means for rapidly decreasing the CO_2 tension of saliva and the increasing salivary pH. Since the change in salivary pH is much smaller than the pH change in the arterial blood, it is obvious that some other factor is also involved in determining the salivary pH" (p. 47). The other factors, according to Brassfield, are the rate of secretion and the type of metabolism in the gland.

Although this evidence may appear to vitiate the hypothesis that salivary pH does not follow blood pH, actually it does not do so, for the typical finding has been that salivary pH and blood pH do not parallel one another until drastic changes have been made in the blood pH. Apparently, a restatement of the position is necessary. Salivary pH will follow blood pH only when extensive or long-sustained changes occur in the blood. With lesser or phasic changes, salivary pH may even shift in the opposite direction, depending upon other effects of the particular substance eliciting the change.

FATIGUE

Since Starr (51) reported a lowering of pH in subjects whose energy level was seemingly lowered, a number of experimenters have described similar findings. Henderson and Millet (27) recorded a drop of pH in one fatigued subject; and Baker and Eye (1), measuring pH electrometrically in four subjects, also noted a drop during rest after exercise. Winsor and Korchin (60) tested several subjects electrometrically and found a drop in pH after exercise, and Nekrasov and Khranilova (39) also reported similar results. These data are in accord with the experiments on HCl injections and the expectations with regard to acidosis.

OTHER PHYSIOLOGICAL MEASURES

Korchin and Winsor (32) tested the rate of salivary flow in 123 subjects aged 12–49 and found that it was not related to height, age, pulse rate, or systolic or diastolic blood pressure. Since the same authors found a direct relationship between rate of flow and pH, it might be inferred that these factors are not related to the pH of saliva. Deakins *et al.* (17) also found no correlation between rate of flow and the age of their subjects. Wenger (58) found that, for 488 aviation students, salivary pH had no correlation greater than .15 with any of 24 other physiological measures except salivary output; however, in a similar study of 201 *Operational Fatigue* patients, Wenger (58) found correlations of .28 with palmar conductance, −.20 with sublingual temperature, and −.20 with O_2 consumption. In a factor analysis of this latter study, salivary pH showed the highest correlation (.48) with a factor expressing the autonomic nervous system.

"ORGANIC" DYSFUNCTION

Stengel (53) determined colorimetrically the pH of saliva in patients with different gastric disorders and found that the pH was not significantly different in subjects with gastric hyper- or hypoacidity. Reindel (42) tested the pH of 50 subjects colorimetrically. In 88 per cent of the cases a "normal" gastric acidity was accompanied by a "normal" salivary pH. Of 22 cases of hypoacidity, 77 per cent exhibited a decreased salivary pH. In the group with hyperacidity, 73 per cent had an increased salivary pH. This would indicate an inverse relationship between salivary and gastric acidity in patients with gastric disorders; however, Wenger (58) found that the salivary pH of 21 peptic ulcer patients did not vary significantly from a group of 488 aviation cadets.

Mayr (37) found that the salivary pH of 25 psoriasis patients ranged between 6.9 and 8.4, with most of the determinations on the alkaline side, while the range for eczema patients was 6.3 to 7.4, with most of the subjects approaching normality. Wenger (58) found no significant deviations in salivary pH in 16 asthma patients.

BEHAVIOR

Among the earliest studies of the changes of salivary pH with behavior was Starr's (51) investigation of four subjects under the influence of emotional stimuli. Secreted saliva tended to become more alkaline during emotional behavior, with an average increase of about 0.4 pH. In the same report he notes that 228 normal subjects had a mean salivary pH of 6.8, while 10 "hyperexcitable psychopathic" stammerers had a mean salivary pH of 7.7. Starr relates this finding to the increased depth and rate of breathing, since 200 specimens from 58 "sub-breathing" stammerers showed a mean pH of 6.0. Hafford (26), who used an electric pH meter, failed to corroborate Starr's findings. She found no significant difference between the acidity of 19 stutterers and normals. The discrepancy in her results and Starr's might have been due to her use of more rigidly controlled experimental conditions and a more precise measuring device.

Rich (44) correlated the salivary pH obtained electrometrically from 39 undergraduate students, with personality ratings made by five of their fraternity brothers, and the salivary pH of 18 graduate students, with personality ratings made by each other and by two of their instructors. The correlations were:

	Good Naturedness	*Perseverance*	*Leadership*	*Aggressiveness*	*Excitability*
39 Undergraduates	.03	.07	−.09	.09	.28
18 Graduates	.09	.24	.10	.51	.45

Rich concludes, with Starr, that the more alkaline is the saliva, the more excitable is the individual. Wenger (58), however, found no significant correlation between salivary pH and the 13 personality factors of the Guilford and the Guilford-Martin scales. His samples of 325 and 197 aviation cadets were rather highly selected and do not preclude that sigificant relationships may exist in an unselected sample.

Winsor and Korchin (60), in a carefully controlled study with an electric pH meter, found that anxiety, fear, and states of depression are accompanied by a decreased salivary pH, and this change is directly related to salivary flow. Since the rate of secretion is easier to determine than acidity, they suggest that the former is a more useful indicator of autonomic balance. Wenger (58), in a study of 225 *Operational Fatigue* patients, in which pH measurements were taken with nitrazene paper, obtained a difference (C.R., 2.50) between the salivary pH of *Operational Fatigue* patients and a group of 488 aviation cadets. The patients had the more acid saliva. Their mean value was 7.1, while that of the cadets was 7.2. Further, a group of 98 patients classified as *Mixed Psychoneurotics* and *Anxiety Neurosis* demonstrated a mean of 6.9 and also differed significantly from the cadet control group. In a later study, Wenger, Furchtgott, and Voas (59), using a glass electrode pH meter, found that a group of 43 ambulatory hospitalized patients from a psychoneurotic ward had a mean salivary pH of 6.61, while a control group of 69 college students had a mean salivary pH of 6.45. The C.R. of 1.9 approaches significance at the .05 level. The mean pH for nine patient with diagnosis of *Mixed* or *Severe Psychoneurosis* was 6.61, for seven patients with diagnosis of *Anxiety Neurosis* it also was 6.61, for seven patients with diagnosis of *Hysteria* it was 6.62, for nine patients diagnosed as *Psychopathic Personality* it was 6.86, and for a group of 11 miscellaneous cases consisting of epileptics, schizophrenics, and other psychotics it was 6.40. It should also be pointed out that of 32 patients listed above in the first four categories, 23 had salivary pH's which were 1 S.D. above or below the mean of the control group.

Wittkower and Pilz (61) examined 27 persons under hypnosis and 16 in the waking state and found that the various affective states increased or decreased secretion, depending on the individual rather than on the conditions of the stimulating situation. The kind of emotional state (fear, joy, disgust, etc.) had no specific influence on the pH changes produced. The determinations were colorimetric.

From the studies of this section it may be seen that some experimenters report a positive correlation between pH and excitability or disturbed states, others report a negative correlation, and still others find no consistent relationship. No more definite conclusions may be drawn at this time.

SUMMARY AND CONCLUSIONS

The recent literature on salivary pH has been reviewed because of the value that the measure may have as an indicator of (*a*) autonomic functions, (*b*) acid-base balance of the fluid matrix, and (*c*) certain overt behavior patterns.

It was found that the pH of mixed, unstimulated human saliva is between 6.44 and 6.85, with the different glands making different contributions to those values.

The following conclusions concerning methodology were reached:

1. The most adequate measuring device is the glass electrode pH meter' although for some large groups of subjects, colorimetric techniques do not seem to displace the mean values significantly from those obtained electrometrically.

2. Prevention of CO_2 loss from collected saliva does not seem to be as important a factor as was previously believed, providing the salivary pH is measured within five minutes after collection. Toluene is relatively ineffective in preventing CO_2 loss.

The following conclusions were reached in regard to the effect of the experimental conditions:

1. It is necessary to control or correct for the effects of:
    a) seasonal variations and particularly air temperature at the relatively high or low ranges,
    b) diurnal activity cycles,
    c) the sporadic occurrence of exercise, and
    d) the time interval since last ingestion of food.
2. Dietary differences have no apparent effect upon salivary pH.
3. When chemical substances are placed in the oral cavity to stimulate

the flow of saliva, the resulting pH is apparently inversely related to the pH of the stimulating object.

4. Saliva collected under conditions requiring jaw movements results in an increased pH.

5. Stimulation of the salivary glands by intravenous injection of autonomic nervous system stimulants or electric current may produce differential pH effects, depending upon the stimulating agent used and the gland stimulated or from which the saliva is collected.

The following conclusions were reached regarding the relationships among salivary pH and other physiological variables:

1. In mixed and parotid human saliva, salivary pH is directly proportional to rate of flow. There is some evidence to the effect that there may be differences in this respect between glands and/or species.

2. Results from subhuman animal experiments indicate that salivary pH does not reflect the acid-base balance of the blood unless great and/or prolonged changes occur in the blood. Contraindicating this conclusion is the fact that in fatigue and during exercise salivary pH decreases as the total blood acid increases.

3. Salivary pH is not related to the age and height of the subject and with the exception of salivary output yields only small correlations with any one of some 24 other physiological measures. It does, however, correlate highly (+.48) with Wenger's factor representing the autonomic nervous system.

4. There are some indications of a relationship between salivary pH and gastric acidity as well as certain skin disorders, but the data are inconclusive.

The following conclusions were reached regarding the relationships between salivary pH and behavioral variables:

1. Salivary pH differs among normal control groups, *Operational Fatigue* patients, and several groups of psychoneurotic patients. This and other evidence points to an "alkalinity-excitability" hypothesis, but the data are not conclusive, and more precise delineations of the experimental conditions are necessary if salivary pH is to be a useful index of various behavioral states.

BIBLIOGRAPHY

1. Baker, K. H., & Eye, M. G. The reliability of the pH of human mixed saliva as an indicator of physiological changes accompanying behavior. *Amer. J. Psychol.*, 1935, **47**, 222–240.
2. Baxter, H. Variations in the inorganic constitutents of mixed and parotid gland saliva activated by reflex stimulation of the dog. *J. biol. Chem.*, 1933, **102**, 203–217.
3. Berg, M., Mayne, A., & Petersen, W. F. Variability of blood pH and its association with meteorological factors. *Amer. J. Physiol.*, 1940, **130**, 9–21.
4. Benetato, G., & Oprisiu, C. Sur les variations du pH et des bicarbonates salivaires dans les états d'acidose et d'alcalose provoqués. *C. R. Soc. Biol. Paris*, 1938, **128**, 113–116.
5. Bienka, E. J. Über Änderung in der Menge der Chloride und der festen Bestandteile sowie über die Speichelsecretion der Submaxillardrüse in

Abhängigkeit von der Art und Stärke des Reizes. *Acta Biol. exp. Varsovie*, 1938, **12**, 93–95.

6. BIENKA, E. J., & SCZEPANSKY, C. Influence de la nature et de la force de l'excitant sur la quantité et sur la teneur en chlorures et le pH de la sécrétion salivaire. *C. R. Soc. Biol. Paris*, 1936, **123**, 32–34.
7. BILLMANN, E. L'électrode à quinhydrone et ses applications. *Bull. Soc. chim. Fr.*, 1927, **41**, 213–280.
8. BRASSFIELD, C. R. Flow and pH changes of submaxillary saliva associated with variations in acid-base equilibrium. *Amer. J. Physiol.*, 1945, **144**, 43–49.
9. BRASSFIELD, C. R., & HANDS, A. P. Changes in pH and the rate of flow of saliva accompanying pH changes in arterial blood during acetylcholine stimulation of the submaxillary gland. *Amer. J. Physiol.*, 1941, **133**, 222.
10. BRASSFIELD, C. R., HANSEN, E. T., & GESELL, R. Acid effects of ammonium compounds. *Fed. Proc. Amer. Soc. exp. Biol.*, 1946, **5**, 10.
11. BRASSFIELD, C. R., & HONG, C. J. Changes in pH and rate of flow accompanying pH changes in arterial blood during chorda tympani stimulation and during pilocarpine stimulation. *Amer. J. Physiol.*, 1939, **126**, 442.
12. BRAWLEY, R. E. Studies of the pH of normal resting saliva. I. Variations with age and sex. *J. dent. Res.*, 1935, **15**, 55–77.
13. BRONK, D. W., & GESELL, R. Electrical conductivity, electrical potential and pH concentration measurements on the submaxillary gland of the dog recorded with continuous photographic methods. *Amer. J. Physiol.*, 1926, **77**, 570–589.
14. CARLSON, V. R., & MCKINSTRY, M. L. Some studies on saliva with particular reference to induced acidity and alkalinity. *Dent. Cosmos*, 1924, **66**, 840–849.
15. CLAFF, C. L., & SWENSON, O. Micro glass electrode technique for determination of hydrogen ion activity of blood and other biological fluids. *J. biol. Chem.*, 1944, **152**, 519–522.
16. CLARK, G. W., SHELL, J. S., JOSEPHSON, J. B., & STOCKLE, M. E. Influence of diet upon inorganic constituents of human saliva. *Dent. Cosmos*, 1927, **69**, 500–513.
17. DEAKINS, M., CHEYNE, V. D., BIBBY, B. G., & VAN KESTEREN, M. Significance of salivary analysis upon small groups of subjects. *J. dent. Res.*, 1940, **20**, 161–170.
18. DE BEER, E. J., & WILSON, D. W. The inorganic composition of the parotid saliva of the dog and its relationship to the composition of the serum. *J. biol. Chem.*, 1932, **95**, 671–685.
19. EISENBRANDT, L. L. Variations in the pH of saliva of five individuals. *J. dent. Res.*, 1943, **22**, 147–156.
20. EISENBRANDT, L. L. Studies on the pH of saliva. *J. dent. Res.*, 1944, **23**, 373–374.
21. GANS, L. R. Experimental study in salivary reactions. *J. Amer. dent. Ass.*, 1926, **13**, 222–225.
22. GASSKILL, H. V. A method for preserving small samples of saliva for subsequent pH determination. *J. gen. Psychol.*, 1933, **9**, 256–257.
23. GLAUBIGER, A., & GALK, K. G. The determination of pH values of biologic body fluids. *J. lab. clin. Med.*, 1940, **4**, 369–376.
24. GREGERSEN, M. I., & INGALLS, E. N. The influence of rate of secretion on the concentration of K and Na in dogs' submaxillary saliva. *Amer. J. Physiol.*, 1931, **98**, 441–446.
25. GROSSMAN, L. I., & BRICKMANN, B. M. Comparison of diurnal and noctural pH values of saliva. *J. dent. Res.*, 1937, **16**, 179–182.
26. HAFFORD, J. A comparative study

of the salivary pH of the normal speaker and stutterer. *J. Speech Disorders*, 1941, **6**, 173–183.

27. Henderson, M., & Millet, J. A. P. On the hydrogen ion determination of normal saliva. *J. biol. Chem.*, 1927, **75**, 559–566.
28. Holzlöhner, E., & Airapetianz, E. Die Drüsentätigkeit bei Nervenreizung. III. Die Wirkung der Sympathikusreizung auf die Glandula submaxillaris und ihr Einfluss auf den Chordaeffect. *Z. Biol.*, 1933, **93**, 491–506.
29. Hug, E., & Marenzi, A. D. Réactions paradoxales de quelques sécrétions après une injection endoveineuse d'acide. *C. R. Soc. Biol. Paris*, 1928, **99**, 240–241.
30. Kesztyüs, L., & Martin, J. Über den Einfluss von Chorda- und Sympathikusreizung auf die Zusammensetzung des Submaxillarspeichels. *Pflüg. arch. ges. Physiol.*, 1937, **239**, 408–418.
31. Kesztyüs, L., & Martin, J. [Physiochemical constants of submaxillary saliva obtained by chorda and sympathetic stimulation.] *Magyar Orvosi Arch.*, 1939, **40**, 346–353.
32. Korchin, B., & Winsor, A. L. The relationship of certain organic factors to individual differences in human parotid secretory rate. *J. exp. Psychol.*, 1940, **27**, 192–194.
33. Krassusky, V. K., Krymskaya, M. L., & Kotlyarevskaya, E. J. [Several features of the parotid gland secretion in ruminants.] *Fiziol. Zh. S.S.S.R.*, 1940, **28**, 372–383 (*Biol. Abstr.*, 1943, **17**, 8380).
34. Lashley, K. S. Reflex secretion of the human parotid gland. *J. exp. Psychol.*, 1916, **1**, 461–493.
35. Leung, S. W. Changes in the salivary volume and pH with changes in oxygen content of inspired air. *J. dent. Res.*, 1946, **25**, 170.
36. McClelland, J. R. The influence of various stimuli upon human saliva. *Amer. J. Physiol.*, 1922, **63**, 127–141.
37. Mayr, J. K. Zur Pathologie der Psoriasis vulgaris. *Arch. Dermatol. Syph.*, 1930, **159**, 598–604.
38. Moore, A. U., & Marcuse, F. L. Salivary, cardiac and motor indices of conditioning in two sows. *J. comp. Psychol.*, 1945, **38**, 1–16.
39. Nekrasov, P. A., & Khranilova, N. V. [Effect of physical work on the unconditioned salivary reflex in man.] *Arkh. biol. Nauk*, 1934, **34**, 593–604 (*Biol. Abstr.*, 1937, **11**, 11074).
40. Nims, L. F. Glass electrodes and apparatus for direct recording of pH in vivo. *Yale J. Biol. Med.*, 1938, **10**, 241–246.
41. Popow, N. A., & Kudrjawzew, W. Über den Einfluss der Reaction der Erregerung der Speichelabsonderung auf die aktive Reaction des Speichels. *Pflüg. arch. ges. Physiol.*, 1930, **224**, 69–74.
42. Reindel, W. Über die Gesetzmässige Gegensätzlichkeit zwischen den Acidverhältnissen von Mundspeichel und Magensaft. *Klin. Wschr.*, 1940, **19**, 390–392.
43. Rich, G. J. The reaction of human mixed saliva. *Quart. J. exp. Physiol.*, 1927, **17**, 53–56.
44. Rich, G. J. A biochemical approach to the study of personality. *J. abnorm. soc. Psychol.*, 1928, **23**, 158–175.
45. Schmidt-Nielsen, B. Microdetermination of pH in saliva. *Acta Physiol., Scand.*, 1946, **11**, 97–103.
46. Schmidt-Nielsen, B. The pH in parotid and mandibular saliva. *Acta Physiol., Scand.*, 1946, **11**, 104–110.
47. Smith, G. H. Factors affecting the deposition of dental calculus. *Aust. J. exp. biol. med. Sci.*, 1930, **7**, 45–77.
48. Soyenkoff, B. C., & Hinch, C. F. The measurement of pH and acid neutralizing power of saliva. *J. biol. Chem.*, 1939, **109**, 467–475.

49. STARR, H. E. Studies of human mixed saliva. I. The determination of the hydrogen ion concentration of human mixed saliva. *J. biol. Chem.*, 1922, **54**, 43–54.

50. STARR, H. E. Studies of human mixed saliva. II. Variations in the hydrogen ion concentration of human mixed saliva. *J. biol. Chem.*, 1922, **54**, 55–64.

51. STARR, H. E. The hydrogen ion concentration of the mixed saliva considered as an index of fatigue and emotional excitation and applied to a study of the metabolic etiology of stammering. *Amer. J. Psychol.*, 1922, **33**, 394–418.

52. STAVRAKY, G. W. Mechanism of synergistic action of pilocarpine and adrenaline on salivary secretion. *Rev. Canad. Biol.*, 1942, **1**, 64–71.

53. STENGEL, F. Vergleichende Untersuchung über die Wasserstoffienenkonzentration (pH-Wert) des Speichels und des Magensaftes bei Magenkranken. *Arch. Verdauungs Kr.*, 1938, **63**, 191–195.

54. STEPHAN, R. M. Changes in hydrogen ion concentration on tooth surface. *J. Amer. dent. Ass.*, 1940, **27**, 718–723.

55. WAINWRIGHT, W. W., & BECKS, H. Human saliva. XVII. Is there a seasonal influence on rate of flow and Ca and P content of resting saliva? *J. dent. Res.*, 1946, **25**, 285–291.

56. WESSINGER, G. D. Comparative measurements of salivary pH. *J. dent. Res.*, 1947, **26**, 123–127.

57. WENGER, M. A. Seasonal variation in some physiologic variables. *J. lab. clin. Med.*, 1943, **28**, 1101–1108.

58. WENGER, M. A. Studies of autonomic balance in Army Air Forces personnel. *Comp. Psychol. Monogr.*, **19**, No. 4.

59. WENGER, M. A., FURCHTGOTT, E., & VOAS, R. B. A study of salivary output and pH in several abnormal groups. To be published.

60. WINSOR, A. L., & KORCHIN, B. The effect of different types of stimulation upon the pH of human parotid secretion. *J. exp. Psychol.*, 1938, **23**, 62–79.

61. WITTKOWER, E., & PILZ, W. Über affektive-somatische Veränderungen; zur affektiven Beeinflussbarkeit der Speichelsekretion. *Klin. Wschr.*, 1932, **11**, 718–719.

62. ZOETHOUT, W. D., & TUTTLE, W. W. *Textbook of physiology.* St. Louis: C. V. Mosby Co., 1943.

Received July 1, 1950.

RESEARCH WITH THE WECHSLER-BELLEVUE TEST: 1945–1950[1]

ALBERT I. RABIN AND WILSON H. GUERTIN
Michigan State College

Since the 1945 review (**111**) of work done with the Wechsler-Bellevue test and since its supplementation (**136**), a vast number of studies have been reported in the literature. Some of them are of major significance, while others have "nibbled" at various problems using the test as one of their main instruments. These past six years have seen a veritable flood of published material with numerous, frequently contradictory conclusions regarding the test itself and regarding its use in health and disease. Partial reviews and summaries concerned with its application to specific problems such as the study of personality (**1**), old age (**95**), and scatter (**100**) have appeared. However, no over-all review and stocktaking account of the whole gamut of applications of the Wechsler-Bellevue test has been undertaken. It is the purpose of the present review to give a general summary of recent research with the W-B[2] test. An effort has been made to cover the literature and investigations in which the test figures prominently. Our concern is with the original scale and not with other forms that have been published subsequently.

Our attempt will, in many instances, question the fruitfulness of the research and the conclusiveness of the findings. We hope that this critical examination will help prevent duplication and repetition of experimentation in certain barren and overworked fields and that it will point out some worth-while areas of investigation.

The first part of the present review will deal with the W-B (Wechsler-Bellevue) as a test of general intelligence, for which purpose it was originally designed. Test modifications, validation, and shorter forms will be discussed under this general heading. The second part covers studies of scatter and subtest patterning in which the main concern is the diagnosis of various and sundry emotional, psychiatric, and neurological conditions.

---

[1] Through September, 1950.

[2] In addition to the Wechsler-Bellevue abbreviation, the names of the subtests will appear in abbreviated form throughout the paper. The single letters *I*, *C*, *A*, *D*, *S*, and *V* stand for the verbal subtests: Information, Comprehension, Arithmetic, Digits, Similarities, and Vocabulary, respectively. The two letter combinations *PA*, *PC*, *OA*, *BD*, and *DS* correspond to the following subtests: Picture Arrangement, Picture Completion, Object Assembly, Block Design, and Digit Symbol. *FS*, *VS*, and *PS* stand for Full Scale, Verbal Scale, and Performance Scale, respectively.

THE WECHSLER-BELLEVUE AS A TEST OF GENERAL INTELLIGENCE

Studies of Reliability

To date, the number of studies dealing with the reliability of the W-B as a psychometric instrument is rather limited. Many investigations deal with retest findings after therapeutic intervention. This material logically belongs to the second part of the present review, dealing with diagnostic problems in psychiatric populations.

Normal subjects. The study by Derner, Aborn, and Canter (2) is the only one dealing with the reliability of the W-B in normal subjects. They administered the test to 158 normals distributed over a fairly wide intelligence range. Sixty subjects were retested one week after the original testing, 60, four weeks later, and the remaining 38, after a period of six months. The mean IQ's as well as mean subtest scores clearly showed practice effects upon retest. The differences were reported to be statistically significant. The practice effect is lessened as the period between tests is lengthened. Coefficients of correlation for the combined groups range from .62 to .88 for the subtests, and the highest coefficient, .90, was reported for the Full Scale. The timed subtests and those subtests involving simple learning showed the greatest practice effects. The authors conclude that, "particularly the Full Scale, can be considered diagnostically dependable with normal subjects" (2, p. 178).

Psychiatric subjects. A similar study reporting retest findings with 53 psychiatric patients was reported by Hamister (61). Thirty-three of the patients were re-examined after an interval of one week, while the remaining 20 were retested after one month. The mean IQ increase for the entire group was 10.25 (statistically significant). No statistically significant differences were obtained between the IQ's of the group retested at the end of one week as compared with the group re-examined at the end of one month. The test-retest correlations of the various subtests ranged from the .50's to the .90's. A different approach, i.e., testing the split-half reliabilities of W-B subtests, was reported by Gilhooly (52). Only four subtests (*I*, *C*, *S*, and *V*) were studied by means of results obtained from 122 neuropsychiatric patients. The reliability coefficients for the subtests were reported as follows: Information (*I*), .79; Comprehension (*C*), .56; Similarities (*S*), .77; and Vocabulary (*V*), .94. The values reported in this study are similar to those reported in the retest studies discussed above (2, 61).

Summary. Thus far, only one study concerning the reliability of the W-B with normals is available in the literature. Studies utilizing psychiatric subjects are too inconclusive because of the constantly changing condition of the patients. On the whole, the Full Scale W-B proved to be a reliable intelligence scale. The reliability of individual subtests varies considerably. Practice effects upon retest have been definitely demonstrated. It has also been shown that the effects of practice depend upon recency of the re-examination.

Correlation with Other Tests

Wechsler-Bellevue II. Gibby (49) found good agreement between Forms I and II of the W-B with a group of 32 neurotics. The correlation between the *FS*'s was .87. The tests were not wholly comparable, however, since comparisons of subtests revealed differences. Comprehension measures something quite different on the two forms as shown by the low correlation of .20 between the subtests. Picture Arrangement was next lowest, with a correlation of .49, while high correlations of .87 were obtained for *PC* and *BD*. A recent study by Hays (63) is concerned with the agreement in patterning of the mental defective on the two forms of the W-B. He found that there are differences in four subtests which were significant at the 1 per cent level of confidence. He did not present correlations as Gibby did, but he found significantly different means for *A*, *V*, *PA*, and *OA* weighted scores on the two forms. Wechsler's profile for defectives was best supported by the W-B Form I.

These studies urge caution in interchanging tests in research studies where something more than an aggregate measure of intelligence is required. It would seem particularly important to perform further investigations of profile patterning on the two tests for the various psychiatric groups, as Hays has done for mental defectives.

Stanford-Binet. Gothberg (56) found the W-B to yield IQ's about 15 points higher than the 1917 revision of the Stanford-Binet (S-B). Correlations between the S-B and the three W-B scales ranged from .73 to .89 with a group of dull-normal subjects. Superior adolescents were studied by Goldfarb (54) with the 1937 S-B. He found Form L yielded IQ scores about five points higher than the W-B. Correlations were .80, .67, and .86 with the *VS*, *PS*, and *FS*, respectively. Sartain (121) also found the 1937 S-B to be significantly higher than the W-B. Correlations of the W-B *FS*, Otis S-A, S-B, and ACE with grade-point averages failed to reveal significant differences. Guertin (57) found the W-B *FS* to give scores approximately five IQ points higher than the 1937 S-B. Subjects were 15 to 22 years old and of high-grade defective to dull-normal intelligence. Kutash (85) used a similar population except that many subjects were older. He found the W-B *FS* was 11 points higher than the 1937 S-B, but he did not control age. He concluded that the magnitude of the differences between the tests was largely a function of this uncontrolled variable. This is hardly surprising, since Wechsler's standardization procedure makes this result inevitable. Reported correlations are attenuated by the uncontrolled age factor and, hence, are of little meaning. Stacey and Levin (131) made a comprehensive report of a comparison of S-B and W-B scores for several groups of defectives. Correlations are attenuated by a somewhat limited range of talent. The brighter subjects had slightly lower W-B than S-B IQ's, while the duller ones had higher W-B IQ's. Tagiuri (**133**) made a comparison of S-B and W-B vocabularies, but the paper was not seen by the authors.

Integrating these studies, we see that the W-B gives somewhat higher IQ's than the 1937 S-B in the lower portion of the intelligence range, but lower IQ's in the upper range. The 1917 S-B is probably still lower than the 1937 revision throughout the whole intelligence range.

Other intelligence tests. Altus (**7**), in his comparison of the Army Wechsler Form B and the W-B, concludes that the latter is slightly more valid. This conclusion seems to be based upon the slight difference between the biserial correlations between acceptance-rejection and IQ's at an Army training center. These were .57 for the W-B and .55 for the Army form. Following Wechsler's point-scoring method, Lindner and Gurvitz (**94**) restandardized the Revised Beta Examination. The correlation between their revision of the Beta and the W-B was .92. In spite of the difference in test content it would appear that they measure essentially the same thing, since the correlation coefficient is of the order of magnitude of a reliability coefficient.

The California Test of Mental Maturity (CTMM) was compared with the W-B by Clark (**20**). The mean total IQ's were comparable for his 50 veterans. The correlation between IQ's was .81. While these tests appeared to be equivalent, a surprisingly large discrepancy was found between the two verbal and the two nonverbal IQ's. Hermann and Hackman (**64**) reported the results of testing two samples of veterans, one with the W-B and one with the CTMM. The IQ's were higher for the sample receiving the CTMM, but sampling biases, recognized by the authors, make conclusions impossible.

The Kent EGY, and W-B were compared by Wright, MacPhee, and Cummings (**143**). The obtained correlation of .51 is rather meaningless because of the extremely restricted range of intelligence. While the mean IQ's are comparable, the authors were astute enough to study the distributions of IQ's in more detail. They found that while dull normals were underestimated, the defectives were overestimated by the EGY. The authors concluded that the EGY is not a satisfactory measure of intelligence for their population. A short series of items was drawn from the EGY for study by Knott and Cannicott (**80**). The psychiatric population showed a correlation of .86 between the EGY and the W-B. Altus (**8**), also interested in a brief verbal measure of intelligence, selected 13 items of general information from a larger battery. The correlation of his battery of items with the W-B *FS* (.73) is rather low, but he seemed to feel that the general information items could be used as a quick verbal measure of general intelligence.

Wright (**144**) studied the Shipley-Hartford Retreat Scale over a wide range of intelligence in a neuropsychiatric population. The correlation with the W-B was .77. Further study of this test by Garfield (**41**) utilized a narrow sample of intelligence, viz., borderline defectives. The subjects were not adequately discriminated by the Shipley-Hartford Scale and the author omits correlational data with the W-B from his report. Magaret and Simpson (**99**) mention incidentally that the correlation between the W-B and the Shipley-Hartford was .87 for 25 heterogeneous hospitalized cases. French and Hunt (**40**) compared the Navy-Northwestern Successive Matrices with the W-B *C-V-S* as the criterion of intelligence. The low correlations between these two measures suggest that the Matrices sample some ability or abilities not measured by the *C-V-S*.

Miscellaneous instruments. Diamond (**31**) presents a report of an interesting investigation of the vocational guidance value of the W-B test. The W-B subtests and various aptitude tests (O'Rourke Survey of Vocabulary, Minnesota Spatial Relations, Minnesota Clerical) were factor analyzed. The subtests most heavily weighted with the three factors are Linguistic: *I*, *C*, *S*; Clerical: *DS*, *D*, *A*; Spatial: *PC*, *OA*, *BD*. He concludes that the W-B is a less adequate measure of vocational aptitudes than other more conventional tests, but that the information from the W-B, when treated in this three-factor frame of reference, is of value to the guidance counselor.

Buck (**17**), who devised the ingenious H-T-P projective drawing test, compared his quantitative scoring with the W-B as a measure of intelligence. He expresses the belief that it may tap aspects of intelligence not usually sampled by conventional intelligence tests. He obtained fair correlations with the W-B, but his sample had an excessive range of talent. His failure to allow for intellectual deterioration with increasing age complicates the comparison. This would tend to attenuate his correlation coefficients and to make the W-B mean IQ too high. Sloan and Guertin (**127**) used the W-B as the criterion of intelligence in their study of the psychometric value of the H-T-P. The limited range of talent employed—borderlines and high-grade defectives—makes the reported correlations difficult to interpret. The authors found the H-T-P IQ to be 15 points higher than the W-B *FS* in their adolescent sample, and they question the adequacy of the H-T-P standardization.

Summary. It is rather obvious from the literature that the W-B is a well-accepted measure of intelligence. Furthermore, the standardization is equally well accepted, as demonstrated by the tendency of investigators to use W-B scores as criteria of intelligence in validating newer instruments. This is somewhat in contrast to the tendency seen in the early review (**111**) of checking W-B validity against other instruments such as the S-B. Only two studies (**7, 121**) reviewed here utilize criteria of intelligence other than psychometric tests. Neither study demonstrated the superiority of the W-B, yet both indicated a fairly high degree of validity for it. Only when the discrepancy between IQ's is between the W-B and the S-B, which is based upon an even larger standardization group, would investigators consider the fallibility of the W-B standardization. The psychometrist's faith seems to be well placed in Wechsler's standardization and the test in general. Some question might be raised of the adequacy and meaning of standardization results at the low end of the intelligence distribution. There is evidence (29) that length of institutionalization influences performance on this test. Wechsler seems to have used some new admissions and some already committed to state institutions, but he does not specify the proportion of each. The test should be used for examination of cases similar to those used in the standardization sample to assure meaningful classifications and profile analyses.

Research-oriented clinicians should see a lesson to be learned from this review, namely, the use of a proper range of talent in correlation studies. Many studies (40, 41, 121, 127, 143) have shown rather inconclusive results because the samples were selected without regard for the distribution of intelligence in the *whole* population. Gothberg (56) is to be particularly commended for her recognition of this requirement and her description and comparison of her sample with Wechsler's standardization population. One comparison of test results (143) disclosed important information, concealed by an over-all comparison of means, through an analysis of the forms of the two distributions of IQ's. This methodological approach indicates the need for more thorough analyses of results than is obtained through the use of standard gross statistics.

Short Forms

Since the first short form (*C-A-S*) was introduced by Rabin, many other forms have appeared. The results of the earlier investigations of short forms were so encouraging that many investigators have published since the last review (111). All the modifications using only W-B subtests are summarized in Table I. McNemar's report (102) is too extensive to include in the table and is reviewed briefly in the body of this section.

Combinations of W-B subtests. Gurvitz (58) suggests using only two subtests, *D-PA*, which is a rather radical step from point of view of brevity. Despite the length of this form and the general feeling about the unreliability of *D*, he found a correlation of .90 with the *FS*, with a large group of subjects. Since this was higher than anticipated, he commendably made further investigations with different populations and found confirmation of this high correlation. Geil (17) proposes the use of a longer form, the *C-S-D-BD*, on the basis of a *FS* correlation of .97 with a sizable group of unselected prisoners. He felt that this form was especially useful in screening out organics and psychotics. In their study of various short forms Cohen and Steisel (24) mentioned the diagnostic value of Geil's form but reported that the highest validity was obtained for *C-A-S*. They do not confirm Gurvitz's findings (50) of higher validity for *D-PA* than for *C-A-S*, but their *N* was only 32. However, Springer's report (130) showed a high correlation of .96 for

TABLE I

Validity Coefficients for Various Short Forms

Form	*Investigator*	*N*	*Population*	*Criterion*	*r*
D-PA*	Gurvitz (58)	523	Heterogeneous	WBFS†	.90
D-PA	Patterson (107)	50	Hetero. psychiat.	WBFS	.81
D-PA	Cohen & Steisel (24)	32	Superiors	WBFS	.52
D-PA	Cotzin & Gallagher (26)	154	HGMD & bdrlne	WBFS	.75
C-A	Cummings *et al.* (27)	420	Bdrlne	WBFS	.93
C-A	Patterson (107)	50	Hetero. psychiat.	WBFS	.85
C-A	Hunt *et al.* (69)	528	Normals	NGCT‡	.58
C-A	Cotzin & Gallagher (26)	154	HGMD & bdrlne	WBFS	.53
S-PA	Hunt *et al.* (69)	528	Normals	NGCT	.72
C-A-S	Springer (130)	100	Bdrlne	WBFS	.96
C-A-S	Patterson (107)	50	Hetero. psychiat.	WBFS	.89
C-A-S	Cohen & Steisel (24)	32	Superiors	WBFS	.78
C-A-S	Hunt *et al.* (69)	528	Normals	NGCT	.78
C-A-S	Cotzin & Gallagher (26)	154	HGMD & bdrlne	WBFS	.59
C-V-DS	Patterson (107)	50	Hetero. psychiat.	WBFS	.93
C-V-DS	Patterson (109)	100	Normals	WBFS	.90
C-A-PA	Hunt *et al.* (69)	528	Normals	NGCT	.72
S-D-PA	Hunt *et al.* (69)	528	Normals	NGCT	.75
C-S-DS	Bay & Berks (11)	200	Court cases	WBFS	.93
C-S-D-BD	Geil (47)	250	Random prisoners	WBFS	.97
C-S-D-BD	Cohen & Steisel (24)	32	Superiors	WBFS	.87
C-S-D-BD	Patterson (107)	50	Hetero. psychiat.	WBFS	.93
C-S-D-BD	Cotzin & Gallagher (26)	154	HGMD & bdrlne	WBFS	.84
V-C-BD-PC	Patterson (107)	50	Hetero. psychiat.	WBFS	.96
V-C-BD-PC	Patterson (109)	100	Normals	WBFS	.96
I-PC-PA-DS	Geil (47)	250	Random prisoners	WBFS	.95
I-PC-PA-DS	Patterson (107)	50	Hetero. psychiat.	WBFS	.95
I-PC-PA-DS	Cotzin & Gallagher (26)	154	HGMD & bdrlne	WBFS	.74
V-I-BD-S	Kriegman & Hansen (83)	207	Bdrlne	WBFS	.95

* The letter abbreviations for the Wechsler-Bellevue subtests are as follows, in order of appearance in the table: *D* = Digits; *PA* = Picture Arrangement; *C* = Comprehension; *A* = Arithmetic; *S* = Similarities; *V* = Vocabulary; *DS* = Digit Symbol; *BD* = Block Design; *PC* = Picture Completion and *I* = Information.

† Wechsler-Bellevue Full Scale.

‡ Navy General Classification Test.

the *C-A-S* with 100 cases of suspected low mentality. This coefficient entails the fallacy of correlating the part with the whole, so Springer reports the correlation of the *C-A-S* with the W-B *FS* minus these three subtests. The obtained coefficient of .89 was still surprisingly high. Hunt, Klebanoff, and Mensh (69) utilized a criterion other than the W-B and found for a large group of normals a correlation of .78 between the *C-A-S* and the Navy General Classification Test (NGCT). This is a fairly respectable degree of agreement between the two tests.

Cummings, MacPhee, and Wright (**27**) found a correlation of .93 between their proposed *C-A* form and the W-B *VS* for a large number of Naval recruits of questionable mental ability. Still another modification was proposed by Bay and Berks (**11**), the *C-S-DS*. They found a correlation of .93 with the *FS* for a mixed sample of court cases and college students. They felt that this form is valuable not only because of its high degree of validity, but also because it retains much of the diagnostic value of the *FS*. Patterson (**107**) was interested in several combinations of subtests proposed by previous investigators (**27**, **47**, **58**, **130**). On a group of 50 variously diagnosed subjects he confirmed the high correlations of the short forms with the *FS* found by the original investigators. He concluded that vocabulary should be included and proposed the *V-C-BD-PC* and *V-C-DS*, which have correlations with the *FS* of .96 and .93, respectively. He feels that many of the diagnostic features of the W-B are retained in these batteries. In a later study (**109**) he found very similar degrees of relationship between his two suggested short forms and the *FS*. Patterson concludes that the selection of a short form for use with a given subject should depend upon the level of intelligence being studied. His *V-C-DS* proved best for superior subjects (**109**). Cotzin and Gallagher (**26**) obtained good results with the *C-S-D-BD*, first proposed by Geil (**47**), when applied to mental defectives. This form was found to correlate .84 with the *FS*, while others (**27**, **58**, **107**, **130**) proved less valuable with this population. Kriegman and Hansen (**83**) studied still another combination of subtests, *V-I-S-BD*, and found a correlation of .95 with the *FS* for a large group of borderlines.

McNemar (**102**) has evolved a statistical technique for arriving at correlations between various combinations of subtests and the *FS* for Wechsler's standardization data. This economical process made it possible to study many combinations. He reports the ten forms with highest correlations with *FS* for pairs, triads, quartets, and quintets of subtests. His correlations are high, but lower than those generally reported in the literature. He attributes this to the use of too broad a range of talent in many studies. However, it is the impression of the present reviewers that there was more frequently a restricted range of talent operating in the studies surveyed here. His enthusiasm for the use of Wechsler's standardization sample is understandable, since it guarantees a proper range of talent, but the use of a nonstandardization type of population is more justifiable if the investigator is concerned with developing a short form for use with that special type of case, as is so often the need. Cohen (**23**) reported the application of a short form to four different psychiatric groups, but the paper was not available for review.

Brief forms using some W-B subtests. Hunt, Klebanoff, and Williams (**69**) incorporated the Thurstone modification of the Stanford-Binet

Form L vocabulary in an attempt to retain this subtest in a more practical form. This study is of special interest since it is the only recent one that has employed a criterion other than the W-B *FS*. Their *C-V-S* showed the highest correlation with their criterion, the Navy General Classification Test (NGCT), among all the forms which they have studied. Their large sample of normals gave a correlation of .86 between the tests. This tendency to find substitutes for W-B subtests in the short forms also appears in a recent study by Hunt, French, and Mensh (**70**), in which they propose the possibility of substituting the McGill Verbal Situation Test in place of *C* or *S* in *C-V-S*. Williams, Hunt, and French (**141**) present normative data for their *C-V-S* based upon 1,000 subjects.

Summary. While most of the studies have been chiefly interested in validity correlations, the agreements in IQ's with the *FS* seem to be at a high level. The attention to correlation is pragmatically justifiable since a restandardization or correction for the short form can be made if the relationship to the *FS* warrants it. Necessary corrections or methods of deriving IQ's from the short forms other than simple prorating have been proposed by some of the investigators (**27**, **83**, **141**).

The correlations cited here are surprisingly high in view of the attenuating conditions that seem to exist. The frequently encountered restricted range of talent (small IQ range) and the diminished reliability of psychiatric populations are the cases in point. On the other hand, conditions are often present which tend to raise the coefficients spuriously. Correlating the part with the whole (short form vs. *FS*) tends to raise the coefficient artificially, although only .07 in Springer's study (**130**). Occasionally, combinations of subtests have been selected to take advantage of sampling irregularities. This has not been a serious error, since large *N*'s have usually been employed and the rechecking of results by later investigators with independent samples has confirmed the original reports. This need to avoid capitalizing on sample irregularities through the use of more than one sample was recognized by several of the investigators (**27**, **58**, **69**, **109**).

All studies of short forms agree in several of their conclusions. It is clear that most short forms are sufficiently valid in their measurement of the same thing that the W-B measures, to be used with borderlines with considerable time economy. The application of short forms to superiors is not so well established, but there is some evidence (**24**, **109**) that the proper combination of subtests will give meaningful results. It would seem desirable to select the proper short form for the intelligence level one is concerned with. Franklin (**38**) has shown that caution must be used in applying short forms to populations other than those on which they were standardized. With delinquent Negro boys the short form investigated is more than four IQ points lower than the W-B *FS* IQ. Unfortunately no correlations were reported.

The studies reviewed here often make the point that no short form retains the full diagnostic value of the *FS*. Some (**11**, **47**, **107**) have tried to select the subtests that would provide the most valuable diagnostic clues, but all agree that there is no substitute for the *FS* in a complete case study.

The intensive research in this area has been facilitated by the availability of routine W-B data. Often one gets the idea that the aim of the study reported was to "get another publication" rather than to make a systematic contribution. Since so many short forms have been proposed and validity demonstrated, it would appear that further research in this area should be confined to systematic attacks such as McNemar's recent contribution (**102**). However, there still seems to be a need for a combination of subtests which will yield valid results on a variety of different populations.

Applications with Special Populations

Superiors. The application of the W-B to populations other than the normal standardization population is a very practical problem. Estes (**33**) points out the difficulties encountered in a scatter analysis of the test results of superior adults. The excessively high verbal and vocabulary weighted scores obtained by these subjects introduces an artifact if scatter is measured from vocabulary. In his study, negative vocabulary scatter on *PC* and *DS* are regarded as a consequence both of using this excessively high vocabulary as a reference point, and the ceiling limitations of the subtests.

Mental defectives. The validity of Wechsler's mental defective subtest pattern for institutionalized mental defectives was investigated and confirmed by Cutts and Sloan (**28**). Cotzin and Gallagher (**26**) presented mean weighted scores in their article which confirm the above investigation. With institutionalized defectives, *A* is extremely low, while *OA* is highest, and, in line with this, the *P* IQ is higher than the *V* IQ. Cutts and O'Kelley (**29**) worked on the problem of the effect of institutionalization on the subtest distribution of defectives. They found that the *VS* was depressed; lowest scores were on *I*, *D*, and *A*. They conclude that their patients with long hospitalization fail to develop verbal abilities to the same extent as other defectives.

Aged. Madonick and Solomon (**98**) were interested in W-B findings with the aged. They gave the test to 50 residents of a denominational home and a home caring for formerly successful executives and professionals. In spite of the fact that their sample appears to be biased in the direction of high IQ's, the authors felt the need to explain the high mean IQ's obtained (*V* IQ, 117.1, and *P* IQ, 115.6). They suggest that the nature of the test construction or the environment, which was stimulating because of a communality of interests, may account for the obtained results. A similar study by Chesrow, Wasika, and Reinitz (**19**) utilized only 20 white males. The order of weighted subtest scores was very similar to that found by Madonick and Solomon. They also found that the vocabulary IQ was usually higher than the *FS* IQ. The *V* IQ's and *P* IQ's were comparable.

Minority groups. A comprehensive analysis of the discrimination of the subtests at various levels among mentally-subnormal Negroes was made by MacPhee, Wright, and Cummings (**104**). They found different patterns at different intellectual levels, i.e., the subtests contributed differently to the *FS* score at the different levels. At the low levels, *S* contributed most to the obtained *FS* score, while as the intelligence increased, all tests came to contribute rather equally. On the basis of subtest correlations with the W-B *FS* they were able to conclude that the W-B is a reliable measure for the southern Negro. Franklin (**38**) also found that the subtests, with few exceptions, discriminated effectively between classifications of intellectual level of delinquent Negro boys.

Hodgson (**67**) studied the test performance of delinquent Mexican boys and found that *DS* failed to discriminate between borderlines and defectives, while *S* failed to discriminate between normals and dull-normals. The order of magnitudes of subtest correlations with the whole test was similar to that obtained with native white Americans. Glaser (**53**) investigated the intelligence of Jewish immigrants. Through an ingenious control and alternate directions in Yiddish and German, he demonstrated the cultural difficulties encountered with *I* and *PA*. Digits was harder for the immigrants, probably because of language handicap and anxieties. Surprisingly, the *V* IQ was somewhat higher than the *P* IQ. The thesis is that performance items (*PA* and *PC*) can prove to be greater obstacles for the immigrant than verbal items because of their cultural loading.

Miscellaneous populations. A unique approach to evaluating intelligence in the blind with the verbal scale of the W-B is seen in a report by Levine and Blackburn (**87**). They studied a large number of W-B verbal scales of newly blinded soldiers. Army General Classification Test scores on these men before their blindness occurred show them to be a representative sample of soldiers with respect to intelligence. The correlation between the prior test and the postblindness W-B verbal scale was .83. The authors found that, unlike individuals who had been blinded for a long time, these cases were significantly lower in *D*. This is attributed to inattention and anxiety. They conclude that as far as their sample is concerned, no impairment of intellectual function can be observed as a consequence of blindness.

Naval recruits, under observation for intellectual deficit, were discriminated well by intelligence levels, according to a report by Lewinski (**89**). Each of the verbal subtests "discriminated effectively," according to the author. Schmidt (**123**) examined the intelligence levels of prisoners with the W-B *FS* and group tests. While the group tests indicated subnormal intelligence, the W-B contradicted this. Lewinski (**93**) also reported an analysis of W-B results on 1,000 soldiers. While the report is primarily a consideration of vocabulary, Wechsler's standardization receives support since all three scales have similar means (near 100) and there is only a very slight correlation between CA and IQ.

Summary. The application of the W-B has not been restricted merely to the standardization population. Application to special groups has been of value when the investigators have taken the differences in subjects into account. The internal validation criterion (subtest vs. whole test) has proved a ready means of attack on these problems. Investigations utilizing this method (**38**, **67**, **89**, **104**) might serve as the models for validity studies of the W-B applied to other special samples.

Refinements and Critiques

Order of subtests. According to Wechsler (**138**) the subtests have their items placed in approximate order of difficulty. Practicing clinicians have noticed that some items seem to be misplaced. A paper by Rabin, Davis, and Sanderson (**112**) reports the order of difficulty for the items in six subtests, based upon the performance of 300 normals and confirmed upon 1,000 psychiatric cases. Some differences from Wechsler's orders of difficulty were found for these populations. Jastak (**72**) has recently reported a similar study with a large group of psychiatric cases. He demonstrated that differences in the kinds of subjects in the sample influence the order of difficulty of the subtests. Some of the sex differences in order of difficulty are especially interesting. In general, his results are quite at variance with Wechsler's data, and he finds that the orders of difficulty are more similar among his various subgroups than between Wechsler's orders and any of those for his subgroups or combined subjects. Certain advantages to an even progression are mentioned in the paper: (1) economy and ease of administration, (2) setting reasonable limits of individual testing, and (3) deriving clinically meaningful measures of intratest scatter.

Sorsby (**128**) presented a revised order of difficulty for vocabulary in connection with a simplified administration procedure. The basal and ceiling levels are established on the basis of only three consecutive successes and failures. He found that these middle 60 per cent of the words were as discriminating as the full range and reduced administration and scoring time by about 50 per cent.

Klugman (**79**) was interested in the effect of order of subtests on digits score. He found that the mean digit score was highest when the subtest came in the middle of the battery. Since the score was higher at the beginning than at the end of the battery, it would seem that fatigue is a more important factor in reducing efficiency on digit recall than is initial anxiety associated with adjustment to the test situation. No psychometric advantage to placement in the middle is demonstrated by this study, and it is of theoretical interest only. No matter where a subtest is placed, it is standardized in that position, so that any constant decrement related to subtest order becomes unimportant. Had the author studied the variances under the three conditions of placement within the battery, he might have been able to attribute greater variance to increased unreliability, which would have practical importance.

Administration procedure. Sappenfield (**120**) recommends the use of cardboard cutouts or templates, which lift up and leave the *OA* pieces in their proper positions for administration. Derner and Aborn (**30**) make a similar suggestion, but use a board with inserts for the *OA* pieces. They are inserted upside down, and a sliding cover prevents them from falling out. The board is placed upside down before the subject, and the pieces drop on the table in their proper places as the cover is slid back. This procedure prevents the subject from seeing the pieces before he is ready to begin the assembly task. These procedures would have their greatest usefulness with neophytes.

Scoring. A practical suggestion by Welsh (**139**) is the combination of time and accuracy scores in one table. This would reduce computational errors in combining the two credits and avoid the possible omission of time credit. Rubinstein (**119**) proposed a new method of accurately recording *BD* performance. The procedure is too detailed to be adopted by the average clinician but could be useful in complete case studies and in research.

Analysis. Foster (**36**) pointed out an error inherent in most clinical evaluations based upon W-B weighted subtest scores. Since Wechsler's weighted scores are based exclusively upon the standardization results of the 20–24 year olds, they can not be interpreted directly when applied to older age groups. Foster suggests that tables for weighted scores for corresponding age groups be included in the manual in order to make meaningful comparisons of subtest weighted scores possible. Barnett (**10**) elaborates on this idea and suggests that the tables give some indication of the importance of the deviation of the obtained score from the mean score for that CA group. This has been done by this author in his tables of deviation-scores in sigma units. Such tables should be of considerable value to research workers who are studying scatter and profile analysis.

Gilhooly (**50, 51**) tested the hypothesis that mean intratest scatter vs. mean score is a negatively accelerated function in the average to superior intelligence range. Only a slight curvilinear relationship was found with his group of 122 neuropsychiatric cases. The *eta* throughout the whole intelligence range was only .25 and was not statistically significant.

Miscellaneous studies. Gerstein (**48**) presented an interesting procedure for analyzing vocabulary responses. She utilizes Yacorzynski's principle that the "right answers" on *V* are not prima facie evidence of adequate knowledge of the word meaning. Her discussion suggests avenues for further research that might disclose more information relating to the nature of verbal performance on this test.

Spache (**129**) found positive correlations between the quality of the responses to *S* and *C* and the weighted scores for the subtests. Some correlation would, of course, be expected since the quality of the response contributes to the weighted score.

A total of 2,000 reports by 50 psychologists were examined by Luchins (**97**). The practices which evoke criticism are many but special emphasis is placed upon the incorrect application of Terman's norms to W-B results. A critique of W-B weighted scores is offered by Brown (**16**). He discusses the "ceiling-floor" effect operating in the testing of subjects with very high or very low ability. He points out that pattern analysis, scatter, and deterioration measures are inappropriately applied when a subject's ability is beyond the ceiling or floor of the test. This is really a discussion of a generally recognized psychometric principle which is not always observed. It is unfair to imply that the ceiling-floor limitations are the fault of a test. The only legitimate criticism of a test is its narrow range of applicability. In this respect, it would appear that the W-B provides as wide a range of measurement as is practical. It can be objected that Wechsler's norm tables include IQ values beyond the desirable limits, which might prove misleading to naive testers who believe that the listing of IQ values is sufficient sanction for using them.

Summary. Some of the suggested modifications in W-B procedure encountered in this review appear to make contributions meriting immediate adoption. Others appear to be of minor importance and are justifiably omitted from standard procedures. It would seem particularly important to publish tables of deviation scores for each CA level for scatter evaluation (**10**), omit the low tabular IQ values in the conversion tables in Wechsler's tables, and to adopt a new vocabulary procedure (**128**).

THE WECHSLER-BELLEVUE AS AN AID IN PSYCHIATRIC DIAGNOSIS

Retest Findings after Therapeutic Intervention

In an earlier section of the present review, some retest studies with the W-B have been considered. These were retest studies per se without the introduction of known modifying factors in the general condition of the subjects. The following paragraphs will deal with retest studies following fairly radical therapeutic intervention.

Shock treatments. Kessler's (**77**) 20 schizophrenic patients were retested with the W-B following 9–25 electric shock treatments. The interval between tests ranged from one to five months. With the exception of three cases, all patients showed IQ increases, the mean increase for the group being 10.6. Though increases were noted in all subtests, statistically significant changes were noted only in *PC* and *C* and the three IQ's obtainable with the W-B. High correspondence (95 per cent) between increases in IQ and psychiatrists' judgment of improvement was also noted. A less detailed and less well-controlled study by Dubin and Thaler (32) reports retest findings on 11 schizophrenics following an unspecified number of insulin and electric shock treatments. The interval between tests ranged between three and 20 weeks. The mean increase in IQ's is reported to be 17 points. All but two subtests (*A* and *V*) showed a statistically significant increase upon retest. However, no relationship between IQ increase and psychiatric condition was obtained. A similar conclusion regarding the lack of relationship between IQ changes and changes in psychiatric conditions was reached by Rabin (113) in ten electrically treated patients. He also found that a rise in IQ is positively correlated with the length of time that elapsed between the last treatment and the retest. Patients retested within a week following the last treatment showed decrements or very small increments in IQ. Those retested from 28–60 days following the last treatment showed a uniformly high increase in IQ. The results are inconclusive due to a possible high sampling error. Further experimental investigation in this area is needed. Fisher (34) employed W-B Forms I and II, alternately, with a group of electric-shock-treated "ambulatory depressed patients." Little change occurred in the unimproved group, but the improved group showed significant changes in *C*, *S*, *DS*, and the three IQ's. It may be noted that this study has two shortcomings: (1) The equivalence of the two forms of the W-B may be doubted (49), and (2) the possible differential loading of diagnostic categories in the improved and unimproved groups is a factor unaccounted for. Carp (18) reported on retest findings in 42 insulin-treated schizophrenics. The mean interval between tests was 73 days. The retest mean IQ was significantly higher than the one obtained during the first testing situation; 76 per cent of the patients showing a gain in IQ, 19 per cent showed a loss, and the remaining five per cent showed no change. The IQ changes were attributed by the author to retesting with the same instrument, rather than to the effects of the therapy. The loss in 19 per cent of the cases, however, remains unexplained. Like Dubin and Thaler (32) and Rabin (113), Carp was unable to show any relationship between improvement or lack of it, and changes in IQ. Five W-B subtests were part of a battery used by Luborsky (96) in examining 12 psychiatric patients before, during, and after a course of electric shock treatments. The five schizophrenic patients included in the group tended to show a drop in subtest weighted scores during the course of treatment. Following the final treatment a rise in scores was general. No measures of significance were reported.

Prefrontal lobotomy. Retest findings on ten patients who underwent prefrontal lobotomy operations were reported by McCullough (101). The patients were examined one week before the operation, and again eight weeks following it. The only statistically significant gains were noted on *D* and *PA*. Most resistant to change were the *V*, *I*, and *S* subtests. A single case study (**145**) reported a loss of 17 IQ points following the operation.

Summary. Retest studies following therapeutic intervention raise considerable doubt whether the demonstrated gains in W-B scores and IQ's are due to the therapy rather than to practice effects. The study with normals (**2**) strengthens this doubt. Results regarding correspondence between rise in IQ level and improved psychiatric conditions are equivocal. Effects of recency of treatment upon the retest findings bear further investigation. Differential effects of therapeutic intervention upon the several subtest scores have not been clearly established.

The Problem of "Scatter" and Its Ramifications

As was indicated in the earlier review (**111**), the W-B test gave the problem of a scatter a new lease on life and controversy. The results obtained with the Binet were negative or inconclusive at best. The earlier review discussed a number of attempts to use test patterns in diagnosis. At best, group differentiations were attained, but individual diagnosis via subtest configuration still remained only a desideratum.

The Menninger study. At about the time of publication of the last review, there appeared a very extensive and intensive study of the use of scatter in differential diagnosis by Rapaport and his collaborators at the Menninger Clinic (**114**). In more than three hundred pages the W-B was described, and a detailed presentation of several measures of scatter in 271 subjects (19 diagnostic subgroups and one "control" group consisting of 54 patrolmen) was given. The three main measures of scatter used in this project were:

1. *Vocabulary scatter*, consisting of differences between the vocabulary score and scores on the other subtests.
2. *Mean scatter*, consisting of "the difference between any subtest score and the average of all the subtest scores" (excepting *D* and *A*).
3. *Modified mean scatter*, consisting of the differences between a verbal or performance subtest score and the mean of all the other scores of their respective scales.

The task performed by Rapaport and his collaborators is a gigantic one. However, as far as the adequacy of experimental design and conclusiveness of findings are concerned, considerable doubt is cast upon this opus by reviewers and other investigators trying to check their results. A summary of the more pertinent criticisms of the Rapaport study, without reference to extraneous stylistic aspects (**15**), is as follows:

1. The "control" group is really not a control group (**46**). Its inadequacy for comparison with pathological groups stems from failure to control sex, age, education, cultural and socioeconomic status, etc. Age is probably most important in this criticism since scores on the W-B, like those on any other intelligence scale, are subject to variation with the age of the subjects.

2. The 19 clinical groups resulting from a breakdown of 217 cases is quite inadequate, since the number in each group becomes much too small for proper statistical comparison and manipulation (**103**).

3. Another criticism, which is also valid in connection with previous scatter studies, is that the scatter measures may be rough indicators of differences *between groups*, but are not valid or useful in individual diagnosis (**46**).

4. For adequate differential dignosis the whole gamut of psychopathological conditions should have been included (**84**). It would have been especially important to include several categories such as organics, psychopaths, and feebleminded, and to differentiate them, respectively, from schizophrenics, neurotics, and a variety of deteriorated conditions.

5. It also seems doubtful whether the use of such statistical techniques as the *t*-test and chi-square on this type of data is justifiable (**46**).

6. The intermingling of clinical observations with the data and conclusions that are primarily to be drawn from quantification findings is very confusing and emphasizes the lack of validity of the results (**3**, **46**).

It appears that the most valuable contribution made in this study by the Menninger group lies in the realm of qualitative description of the W-B subtests and in the hypotheses made regarding the possible psychological processes involved in these performances. Krugman (**84**) states it well in his review: "Using the data as a springboard, Rapaport has employed a combination of unusual clinical insight, acute analytical treatment and profound theoretical reasoning, to synthesize a diagnostic approach . . . " (p. 285). This statement, probably minus the glowing adjectives, summarizes well the nature of the contribution.

From the viewpoint of quantitative research and experimentation this study has failed to produce any clear-cut or conclusive findings concerning scatter. On the contrary, it has added to the already existing confusion of contradictory findings on various groups which differ in age and in control of the several pertinent factors involved. Some of the studies quoted in the latter part of the paper will have a further, direct bearing upon the methods and conclusions of the Menninger group.

Other measures of scatter. Pascal and Zeaman (**106**) have empirically derived an index of W-B scatter which differs somewhat from previous attempts. It is based on differences between three pairs of subtests (*I-C*; *BD-DS*; *V-PC*) and the absolute difference between *S* and *BD*. The added resulting figures comprise an "index," the application of which shows significant differences between normals and neurotics, and between neurotics and psychotics. Moreover, they conclude that their index reflects "the progressive seriousness of psychological disturbances" and that it gives "an estimate of level of adjustment." Absence of psychological rationale and group, rather than individual, differentiation are the major criticisms of this and many similar approaches. The index relationship to diagnosis may suffer from the same shortcomings of another empirically derived "ratio" criticized by Webb (**137**).

Holzberg and Deane (68) have suggested a measure of "intratest scatter" instead of the conventional intertest scatter within the W-B scales. The scatter of the items within some subtests was reported by the authors to distinguish to a significant degree between organics, neurotics, and schizophrenics. The "altitude" approach to the measurement of scatter was suggested recently by Whiteman (140). This measure utilizes the top scores of subtests, rather than their mean, as a reference point for deviation and scatter analysis. Statistically significant differences between normals and schizophrenics were reported as a result of the application of this type of scatter.

Scatter in differential diagnosis. Using Rapaport's data, Wittenborn (142) tested several neatly arranged hypotheses which are important in profile or scatter diagnosis. The hypotheses may be roughly stated as follows:

1. All patients (Rapaport's) having the same diagnosis have a similar W-B test score pattern.
2. Subtest correlations between pairs of patients of the same diagnostic group are higher than those between pairs of different diagnosis.
3. Correlation coefficients between pairs of patients of the same diagnosis are higher than coefficients obtained "between pairs of hypothetical patients for whom two or three of the subtests have the same rank" (p. 437).

The method of rank order correlation was used in the investigation of each of these hypotheses. The coefficients of correlation (in terms of *rho*) were so small that they cannot be considered as support for the hypotheses. The author concludes that for Rapaport's groups a scatter pattern "is of no greater value as a psychometric supplement to diagnosis than an inspection of the subtests for a few conspicuous instances of failure and relative success" (p. 439).

In his review of scatter analysis, Mayman (100) was rather optimistic about its value in differential diagnosis. However, subsequent work did not seem to justify entirely this optimism. Patterson (108) reported 78 per cent agreement between psychologists' diagnoses based on scatter analysis and that of psychiatric diagnoses is open to several major criticisms. The "blind" diagnoses were not actually blind since the majority were seen by the psychologist who diagnosed them; the test reports were a factor in establishing the psychiatric diagnoses. Moreover, test patterns (scatter) on which the diagnoses were based were not reported. A close examination of the actual comparison of "agreements" by Garfield (42) shows that the level of time agreement is really much lower than the 78 per cent reported. A study of similar design by Teicher and Singer (134) reported 52 per cent agreement between scatter and psychiatric diagnosis. Another 8 per cent showed agreement to some extent. Test patterns of 50 neurotics and 50 schizophrenics, rigidly controlled with respect to a number of significant factors (age, sex, education, etc.), were reported by Kogan (82). In this study pairs of patients having similar W-B patterning were compared with each other, regardless of diagnosis. The author's conclusion is that his "investigation does not support the commonly held conception that the WB patterns are a useful tool in the differential diagnosis of functional disorders . . . " (p. 44). Similar results, prompting "a note of caution," were also reported by Peixotto (110). Inadequacy of finer differentiation within the normal range has also been reported (35).

Summary. It is quite possible that some of the high correlations between psychiatric and W-B diagnoses obtained by different investigators are due to the inspectional techniques of the clinician rather than the numerical scatter values. One may also agree with the reviewers (1) who state that "Score-patterns successfully differentiate groups, but are of comparatively uncertain value in the diagnosis of individuals" (p. 108). The efficacy of W-B scatter in *individual* or differential diagnoses has not been demonstrated. Intratest scatter (68) and "altitude" (140) as a reference point in scatter may merit further investigation.

Indices of Deterioration

A more particular aspect of scatter that has been reported extensively in the literature is that of deterioration. In the third edition of *Measurement of Adult Intelligence*, Wechsler (138) proposes an index of deterioration, which he calls the "Mental Deterioration Index" (MDI).

Diagnosis of organicity. Levi and his collaborators (86) found this method useful in differential diagnosis between psychogenic and organic aspects of disorders in the 15 cases described in their clinical study. However, these investigators seemed fully aware of the inconclusiveness of their findings. They left the following points to be clarified by future investigations:

1. Determination of a critical percentage of loss indicative of definite deterioration.
2. The extent to which psychogenic disorders may simulate organic conditions.
3. The extent to which test deterioration is unaccompanied by physical symptomatology.

The test pattern and the application of the MDI to 50 brain-injured patients were studied by Allen (3, 4). The results shows that the *I* subtest is most resistant to the effects of brain injury and that the *OA* is one of those most affected. Moreover, the application of the MDI disclosed that only 54 per cent of the brain-injured were identified by it (using 20 per cent as the "definite" indication). Allen also suggested a modified index consisting of the difference between the sums of the weighted scores of two pairs of subtests (*I* and *C* minus *D* and *DS*). His suggestion was that a difference of 5 points or more between these pairs of subtests would be a more definite index of deterioration.

Blake and McCarty (**12**) compared Allen's data on the brain-injured with results from 50 college students. Using the 10 per cent level as critical, they discovered that 14 patients and 12 normals were misdiagnosed. Thus, according to Wechsler's method, 26 per cent of the total group of 100 subjects were misdiagnosed. When they applied Allen's modification to the normals, 17 cases, or 34 per cent, were misdiagnosed. The authors admit that all they have shown is some statistical significance for the MDI; however, they caution that " . . . this report is not concerned with the efficiency of this score in differential diagnosis" (p. 418).

Recent work by Kass (**76**) essentially shows agreement with Allen's (**4**, **5**, **6**) findings. He concludes that in two-thirds of his cases the MDI failed to detect or confirm cases of brain damage.

Deterioration versus mental deficiency. Boehm and Sarason (**13**) chose to investigate whether the formula offered by Wechsler is successful in distinguishing intellectual deterioration from mental deficiency. Twenty-two "familial" mental defectives ranging in age from 15–63 were the subjects of this study. In all instances the neurological examination was negative, and feeblemindedness was evident from early childhood. The mean deterioration loss was 15 per cent for the entire group. Fourteen of the cases showed losses from 22 to 48 per cent. The authors, therefore, quite correctly conclude: "The deterioration index . . . cannot be considered as a valid tool for distinguishing between mental deficiency and deterioration" (p. 358).

A study by Sloan (**125**) lends added support to the conclusions of Boehm and Sarason (**13**). This investigation applied Wechsler's MDI to 80 male mental defectives, predominantly of the familial variety. Using the 20 per cent criterion, nearly 70 per cent of this population would be considered as deteriorated. However, neither clinically nor psychometrically could such a contention be supported. Fortunately, intelligence test findings were available on 56 cases that had been examined during previous years. A re-examination did not produce any significant differences. The coefficient resulting from a correlation between the magnitude of the difference and the deterioration index was positive, but too small to be considered statistically significant. Thus, Wechsler's MDI was shown again to be inadequate in differentiating "deterioration" from mental deficiency.

Age and deterioration. Hunt's analysis (**71**) of Wechsler's original standardization data, in which the subtest score means of the young, nondeteriorated group (ages 20–24) are compared with those of an old group (55–59), showed that only two of the subtests seem to "hold up" with age (*I* and *C*). Moreover, the author also shows that *BD* and *DS* are the most vulnerable subtests. He consequently suggests that a ratio based on these pairs of subtests might be most sensitive and distinguishing.

Copple's analysis (25) of 1,403 records of a wide age range (20–69) awards "some validity" to Wechsler's MDI. The author feels, however, that the use of his conversion table for obtaining a senescent decline score, based on a multiple regression equation, is preferable. A study referred to earlier (19) supports the use of the index in the aged and shows a positive correlation with CA.

Johnson (73) correlated the MDI with CA for sizable groups of epileptics and feebleminded. The coefficients of correlation obtained were quite high for two epileptic groups (.60 and .46) and quite low (.22) for the feebleminded. Yet, the correlation between the length of time the patients had had seizures and the MDI is reported to be low by the same investigator. Thus, the conclusion one is tempted to draw from the first part of the study—that there is a relationship between age in epileptics and the MDI, which can probably be attributed to the history of seizures—is not supported in the second part. Since there is considerable difference in the reported mean ages of the groups, the results are difficult to interpret.

Differentiation between various groups. A comparison between psychoneurosis and schizophrenia was made by Schlosser and Kantor (122). Their population of 163 subjects consisted of 56 psychoneurotics, 65 paranoid schizophrenics, and 42 mixed schizophrenics. They did not use all of the subtests in the scale. The "hold" subtests used in the ratio were: *I*, *V*, *PC*, and *OA*. The "don't hold" subtests were: *D*, *A*, *BD*, and *DS*. A priori, the expectancy would be that the schizophrenic group, particularly the nonparanoid, would show a higher loss than the psychoneurotics. However, no statistically significant differences were obtained between the three groups. It appears that when the deterioration index is used, psychoneurotics, paranoid, and mixed schizophrenics show similar levels of loss of efficiency. The inadequacy of *I*, *C*, and *V* as indicators of the level of intelligence in the premorbid state of patients was pointed out by Rappaport and Webb (115).

The MDI was investigated by Rogers (117) with a total of 349 subjects, subdivided into seven groups, some described in the literature and others obtained from counseling and mental hygiene services. The findings indicate that normals differ significantly from the clinical groups. However, the maladjusted groups do not differ significantly from each other when the MDI is used. The index does not seem to differentiate adequately between organics and one group of patients with functional disorder. Results similar to those reported by Blake and McCarty (12) were obtained by Rogers (118) when Allen's index (4) was used.

Comparison with other measures of deterioration. Garfield and Fey (45) studied the relationship between Wechsler's measure of deterioration and that obtained from the Shipley-Hartford. The 100 hospitalized males revealed an agreement of .37 between these two measures.

Magaret and Simpson (**99**) report a correlation of only .10 between these two indices of deterioration. Furthermore, they failed to find evidence for validity of either method of estimating deterioration. The clever procedure of correlating psychometric decline over a period of illness with measures of deterioration, and a comparison of the indices with psychiatrists' ratings of deterioration gave negative results.

Fox and Birren (**37**) compared Wechsler's MDI with findings in the Babcock-Levy test. The correlation obtained between the two indices in 50 senescent normals was not significantly different from zero. The need for independent validating criteria was emphasized by them.

Summary. Wechsler's original remarks (**138**), based on empirical results and not entirely borne out by his own statistics, set off a wave of studies, the results of which are predominantly negative. The original discussion of the deterioration index was chiefly in connection with the "normal process of deterioration" due to age. However, many investigators and clinicians did not lose time in jumping the gap from old age to organicity. It may well be that the reduction of mental functioning due to old age is qualitatively different from brain trauma or brain disease. Moreover, does all brain trauma or disease of necessity produce marked deterioration? Are not the foci of pathology and the pretraumatic and predisease personalities important factors to be considered before all of these cases are tossed into the "organic hopper?"

A more serious indictment of the MDI comes from the studies on familial mental deficiency. They raise the question whether the "index" is only characteristic of old age, or whether it may not be the result of a pattern to be found in a variety of conditions not related to old age. Another difficulty is the inadequacy of the MDI in differentiating "organics" from some other functional groups. Further statistical refinement of the index followed up by its application to a variety of comparable (age, IQ, etc.) clinical populations may produce more definitive results concerning its value as a clinical tool in differential diagnosis.

Effect of Organic Involvement

An attempt to measure changes in the intellectual functioning due to physiological or somatic variations and due to structural brain insult has been made by several investigators who utilized the W-B scales. In these instances, the investigations deal with effects noted in the entire scale or the subtests in it, but not with any specific "deterioration index."

Differential diagnosis. One of the studies which falls in the category described above was made by a group of investigators at the Boston Psychopathic Hospital (**55**). Over a 13-month period, 428 patients were examined by means of the W-B scales. Of these, 92 were considered as showing "patterns" of brain damage. These judgments were based on eight subtests of the W-B scale (exclusive of *OA*, *DS*, and *V*). When checked by clinical examination (history and a variety of neurological tests) 72 per cent of the cases were verified by clinical diagnosis; in 10 per cent the diagnosis was inconclusive, and in the remaining 18 per cent no clinical evidence of brain damage was obtained. Seventeen per cent of 40 hospitalized patients "with no psychological abnormalities to suggest brain damage" had some abnormal clinical findings. The study suggests numerous valuable *qualitative* observations based on the work of Goldstein and his collaborators. However, most of these observations are not, strictly speaking, the W-B test. The authors suggest such qualitative signs as irritability, mood changes, irregularity in speed of performance, stickiness in shifting, circumlocution, slurring, and many others as indications of organicity. It may be questioned to what extent these signs overlap with the "mental examination" on which the clinical diagnosis is based at least in part. The inspection of the figures of subtest distribution yields a "semiquantitative" conclusion that the patients with brain damage "tended to show relatively great preservation of ability on comprehension and information items. . . . Considerable loss . . . on *D*, *A*, and all the nonverbal items" (55, p. 172). This conclusion is based on a general impression rather than on detailed statistical analysis and tests of significance. One cannot agree that these findings "establish" the W-B as a reliable instrument in the diagnosis of brain damage. All that can be said is that some psychologists acquainted with clinical signs of organicity may be able to make a diagnosis of brain damage while using the W-B test.

An interesting approach to the diagnosis of cerebral pathology with the W-B test was presented by Hewson (65). This worker used as many as 13 ratios between various combinations of subtests in a comparison of patients with cerebral pathology (concussion cases, cerebral arteriosclerosis, CNS syphilis, etc.) with a variety of psychoneurotics and a group of normals. These ratios apparently have been empirically derived. The author states that the obtained quotients of the normal group were distributed between those of the neurotics and the cerebral pathology group. However, an examination of whatever meager quantitative data are presented does not justify the statement. It is also interesting to note that this worker advocates combinations of ratios for adequate diagnosis. A great deal of raw data on the 59 neurotics and 98 cerebral pathology cases are given. In addition, some results on normal subjects and on patients having a variety of additional diagnoses were presented. The relationship of these data to the efficacy of the ratios in diagnosis is not clear. For some of the numerical data presented, the indications are that the overlapping between organics, normals, and neurotics is considerable when the combination of the several ratios is used. The connection between some of the speculations and alleged conclusions and the data presented is very nebulous and hardly comprehensible to the present reviewers.

This procedure was also applied by Guttman (59) in comparing 30 organics with 30 normals. Sixty per cent of the organics and 17 per cent of the normals were diagnosed as organic. The author considers Hewson's method superior to other types of W-B indices of organicity.

Epilepsy and migraine. Lewinski's studies on epilepsy (92) and migraine (91) may also be included in this section, though the organic basis for both of these disorders is far from having been established. These two disorders show similar rank ordering of subtests. The *rho* coefficient of correlation between the rank order of subtests in epilepsy and migraine is .936—almost unity. The *rho* when epilepsy is compared with neurosis is lower (.718+.11) but statistically significant. However, no significant correlation between normal and organic subtest ranks was obtained. The group findings show that performance was particularly good on *OA*, *BD*, *PC*, and *C*.

Single case reports. An interesting speculation regarding organicity in test performance was brought forth in Cofer's study (22) of a case of postinfectious encephalopathy in whom definite neurological involvement was found. The W-B (exclusive of *A*, *DS*, and *PA*) was administered on three occasions with a seven- and a two-week interval. The second administration was under hyoscine. Under the effects of the drug, the total IQ rose 14 points. The rise in the performance IQ was much more marked than was the verbal IQ. A lapse to the previous level of performance took place two weeks after the treatment was terminated. Though this is a single case, it is reasonable to question whether the deficit which may be observed in such cases via test performance is stable and unidirectional. Possibly other factors than the brain damage were operating.

Kogan's case report (81) of a meningo-encephalitic patient indicates highest subtest scores on *I* and *A* and particular difficulty with *C* and *BC*. Thinking, as a result of qualitative findings from the W-B and other tests, was considered to be extremely "concrete." These results are fairly consistent with Cofer's findings (22) regarding the rise in the performance tests under hyoscine.

Summary. It may be stated that the studies reviewed in this section add little to our systematic understanding of intellectual changes in organic disorder. The variety of patterns in "organics" raises the question whether the structural insults of different localizations may produce differential effects upon intellectual functioning. The possibility exists that some common elements observed in the patterns may be due to a similar *reaction of the personality* to the damage. It is unfortunate that many of the studies do not report all of the scores for the W-B subtests, thus making precise comparison of findings regarding intra-individual patterning very difficult, if not impossible.

Schizophrenic Patterning

Considerable impetus to subsequent research in schizophrenia was given by the studies reviewed in 1945 (111) and by the publication of

Wechsler's revised edition of *Measurement of Adult Intelligence* (**138**). Wechsler expanded his discussion of diagnostic indications in the test.

Studies utilizing Wechsler's diagnostic indicators. Brecher (**12**) obtained from Wechsler's discussion as many as 14 diagnostic "signs" for schizophrenia, 12 for organics, 13 for neurosis, 13 for psychopathic personality, and 12 for mental deficiency. Her study of a group of 40 schizophrenics indicates the futility of these signs. If the patients who were diagnosed clinically as schizophrenic were to be diagnosed on the basis of Wechsler's signs of schizophrenia as opposed to those characteristic of other disorders, only 32.5 per cent of the cases would be correctly diagnosed.

Olch (**105**) compared test results on 32 schizophrenics with those of two age groups (17–29 and 40–49) of Wechsler's standardization population. A number of characteristics of the schizophrenic pattern suggested by Wechsler and other investigators mentioned in our earlier review (**111**) were critically examined. Olch found no significant differences between the verbal and performance IQ's of schizophrenics; neither did she find that the age factor has any effect on these conclusions. However, she did find significant differences between younger schizophrenics and normals in individual subtest scores. These subtests are: *DS*, *C*, *S*, and *PC*. A finding substantiating earlier studies is the greater variability in test performance of schizophrenics as compared with Wechsler's normals.

In Levine's study (**88**), five judges were asked to diagnose W-B records on the basis of Wechsler's discussion in the recent revision of his book (**138**). The records were those of 110 white males distributed among a large number of diagnostic groupings. Among these, 46 cases were diagnosed as schizophrenic. The findings indicate that: "the agreement between WB diagnoses is of borderline statistical significance. The percentage of agreement ranges from 59.7 to 67.2 for the cases classified" (**88**, p. 29). Despite the fair statistical reliability for the ratings, the disagreement in diagnosing schizophrenics vs. nonschizophrenics ranged from 17–32 per cent, and there was overlapping of the profiles and unreliable discrimination in some cases. Thus, the separation of schizophrenics from other patients on the basis of W-B patterns is only somewhat better than chance, but considerably better than the results obtained by Brecher (**14**) who employed the "sign counting" method in examining their efficacy for *individual* diagnosis.

Johnson (**74**) obtained his score patterns by averaging the individual's subtest deviations from the mean of his subtest scores. Agreement of observed subtest deviation with that proposed by Wechsler ranges from 8 per cent on *OA* to 87 per cent on *S*. Variability was greater than reported by Wechsler, sex differences were not found, and a positive relationship between variability and intellectual level was obtained, *Other schizophrenic patterns.* Garfield (43), in his study of 67 schizophrenics and 46 "controls" (nonpsychotic psychiatric patients), utilized the deviation of each subject's subtest scores from his own mean of the subtests, excluding vocabulary. The data were reported in five degrees of deviation, two positive and two negative, clustering around an insignificant middle category. Table I shows the largest differences in the *BD* test in which the schizophrenics are favored, while the control may be, on the other hand, compared favorably with the schizophrenics on the *C* test. Though other trends were noticeable, there was "considerable overlapping between the schizophrenics and control groups." It also appears to this investigator that "no clear-cut patterns are discernible." The alleged higher verbal score than performance score in schizophrenics is criticized and receives no statistical support. Rabin's index, Rapaport's idea (96) that *BD* is less affected by schizophrenia than *PC*, and Wechsler's "tentative criteria which could be checked objectively" (*OA*, *BD*, *PA*+*PC*, *I*+*BD*) are also criticized and given no support by Garfield's results. The general conclusion of the author is that "there are no reliable scatter patterns on the Wechsler-Bellevue scale for schizophrenic subjects" (43, p. 35). In view of the inadequacy of the "control" sample (neurotics, psychopaths, alcoholics, etc.) it may be wondered whether such a categorical statement is warranted.

In a later study (44) the same investigator reported on a larger sample of 109 schizophrenic patients. In a comparison between the ranks of the means of subtests of his own group with those of Olch (105), Magaret, and Gilliland *et al.*, mentioned in the earlier review (111), he points out the inconsistency between group findings. The *DS* is about the only subtest that shows consistently low scores in the samples compared; other subtests, however, show variations of considerable magnitude. He feels, quite justifiably, that the assumption that different schizophrenic groups are comparable is not warranted, due to the differences in diagnostic criteria used. It also appears that the discrepancy between the verbal and performance scales (in favor of the former) is a function of educational and intellectual level. The difference is considerable in the educationally and intellectually superior group, while it is nonexistent in the average and subaverage groups. Variations in the rank order of the several subtests in the different groups were also noted. The author feels that in view of his findings, Rapaport's (114) relative neglect of such factors as age, IQ, and educational level in diagnostic studies of scatter is not justified. Finally, a tabulation of the percentages of subtest deviation shows wide variability for the entire schizophrenic sample and leads to the varification of the previous conclusion that there is no clear-cut schizophrenic test pattern.

Klein's (78) "application of the multiple regression principle" is illustrated in the study of Rapaport's (114) 72 schizophrenics and 54 normal controls. A comparison of these two groups on vocabulary scatter, employed by Rapaport, yields some interesting results. "Schizophrenic impairment" is due least to *D* and *OA* and is most evident in *C*, *PC*, *A*, and *DS*. In attempting to differentiate *individuals*, the regression equation misidentified 26 per cent. The "linear combination of subtest scores yielded a multiple *R* of .45 with the criterion of group identity" and is significantly different from zero. Subtests *C*, *PC*, *DS*, and *A*, in order of decreasing effectiveness, were almost entirely responsible for "the discrimination capacity of the battery." The critique of the sampling method that may be applied to Rapaport's work is also applicable to this study.

Subtypes of schizophrenia. Freeman (**39**) compared 28 hebephrenic schizophrenics and 28 paranoids with the same number of normals matched for age, sex, race, educational background, and even religious affiliation. The data reveal that the discrepancy between the verbal and performance scale scores markedly favors the verbal score for the paranoid group, but is in the opposite direction in the hebephrenics. The total scores, verbal scores, and performance scores are highest for the normals and lowest for the hebephrenics. The hebephrenics function at a significantly lower level than do the controls on both scales, while the difference between them and the paranoids is only significant with respect to the *VS*. A significant difference between the paranoids and controls was found on the *PS*. The *FS* comparisons between the three groups were all statistically significant. The *FS* scatter and *PS* scatter did not show discrimination between the groups. The only significant difference on verbal scatter was found in the comparison of the hebephrenics and controls. The use of covariance did not alter the findings to any appreciable extent. Freeman concludes that the "scatter approach has not yet proved a valid one in the diagnostic testing of hebephrenic and paranoid schizophrenics."

An attempt to measure patterns in 245 schizophrenics was made by Harper (**62**) by means of Fisher's discriminant function. Reliable group differentiations between some major types of schizophrenia were obtained. However, the final conclusion is that the W-B ". . . is not a highly adequate tool for psychiatric diagnosis" (p. 295).

Summary. The findings concerning the use of the W-B in the diagnosis of schizophrenia are, in the main, negative. The results concerning the discrepancy between the *VS* and *PS* in schizophrenics remain inconclusive, since they differ from study to study. Individual differentiating subtests are not discriminating between schizophrenics and other groups. Several scatter approaches (mean scatter, vocabulary scatter, etc.) do not demonstrate the reliability of a specific schizophrenic pattern. Despite repeated warnings in the literature, many significant factors remain uncontrolled in many of the studies. Wechsler's "sign" approach did not prove fruitful in differential diagnosis. Judges utilizing Wechsler's diagnostic hints were somewhat more successful. Apparently, a combined approach in which some quantitative data are utilized along with "clinical insight" is most practical in the diagnostic process.

Patterns in Neurosis

Anxiety. A scatter analysis (deviation of each subtest from the mean of all subtests) of the W-B records of 25 anxiety neurotics was reported by Lewinski (**90**). No characteristic pattern was obtained beyond a few qualitative observations. Rashkis and Welsh (**116**) devised 12 signs of anxiety which were based on W-B records of 30 cases. Seven of the "signs" are due to "temporary inefficiency" on *D*, *I*, *BD*, *A*, *OA*, *PC*, and *PA*, while the remaining are extraneous and qualitative in nature (apprehension, distractibility, somatic complaints, etc.). A comparison between the anxiety patients and a group of nonanxiety patients did not yield statistically significant differences in the incidence of the signs. The qualitative signs appeared among the ones that tended to differentiate between the groups most effectively. Shoben's comparison (**124**) of 18 anxious with 17 nonanxious college students yielded no significant differences in the number of signs listed by Rashkis and Welsh (**116**).

Warner's (**135**) study reported a comparison of W-B findings in 60 anxiety neurotics with those of 65 normal controls. Excepting the higher relative position of *PA*, *OA*, and *PC* in the anxiety neurotics, the investigator found that intertest variability, differences between verbal and performance IQ's, and relative *D* and *A* scores did not differentiate between the two groups.

A study in which a number of important factors were controlled in a comparison of 50 anxiety reaction cases with 25 orthopedic hospital cases, was reported by Heyer (**66**). A comparison of the groups with respect to mean weighted subtest scores, intertest variability, and subtest deviation from vocabulary yielded no positive results. *Group* results supported Wechsler's contention of *V* IQ higher than *P* IQ. However, this sign proved useless in *individual* diagnosis. No characteristic "pattern" of anxiety was produced on the W-B scales. A question regarding the adequacy of the control group may be raised. Orthopedic surgical patients frequently may have severe enough anxiety reactions to be comparable to those of anxiety neurotics.

Compulsive alcoholism. Similar negative conclusions have been reached by Halpern (**60**), who studied a group of 47 compulsive drinkers. There was no "characteristic organization of mental abilities which would serve to distinguish them either from normal subjects or from other clinical groups" (p. 472).

Summary. The results of the several studies do not support the contention of quantitative W-B findings characteristic of anxiety or compulsive trends. Some helpful clinical and qualitative observations may be made during the administration of the test, but they are not an integral part of it or uniquely evoked by it.

Psychopaths and Delinquents

Strother (**132**) studied the W-B's of 14 psychiatrically diagnosed psychopaths as a check on Wechsler's proposed pattern. Only one sign, low A, was in the expected direction, so the author concludes that these signs do not appear with sufficient consistency to be useful. Institutionalized behavior problem groups were examined by Clark (**21**). His Mexican and American soldier and civilian problem groups all showed depressed W-B VS, low A, high OA, and $PA+OA$ greater than $PC+BD$. He concludes that Wechsler's psychopath pattern (**138**) is substantiated. Franklin (**38**) applied the W-B to delinquent Negro boys. Although aware of the fallacy of comparing their pattern with Wechsler's, which was obtained from a different type of sample, he found considerable agreement. Altus and Clark (**9**) studied a group of white and a group of Mexican institutionalized problem boys. Since four of Wechsler's adolescent psychopath signs were found, the authors concluded that his pattern was corroborated. Sloan and Cutts (**126**) studied institutionalized defective delinquents. They concluded that there is some agreement with both the Wechsler pattern for mental defectives and that for psychopaths. However, a control group of nondelinquents (**28**) showed just as many of the psychopath "signs."

Summary. The notorious lack of consistency in the use of the label means that different investigators study different subjects. The test results then are treated as group data, with means, which will obscure important individual differences. There is also the questionable application of the *adolescent* psychopath pattern of Wechsler to mature subjects. Lastly, it should be mentioned that in the original data (**138**, p. 214, No. 15) from which Wechsler derived his pattern, the "nonpsychopathic" delinquent controls showed a test pattern that was as similar to that of the "psychopaths" as it was to the standardization population. Considering all these difficulties, the reader will probably be surprised to see *any* agreement between investigators.

GENERAL SUMMARY

The eleven years since the original publication of Wechsler's scale of "global intelligence" in 1939 have shown that his instrument is an easily adaptable and pliable tool for research purposes. Recent studies no longer question the validity of this test as a measure of intelligence (**111**). On the contrary, the vast popularity and wide usage of the test have made it a commonly used measuring rod for comparison and validation, if not actual calibration of newer and more recent techniques. Moreover, even parts of the scale or certain combinations of its subtests have been variously reported and advocated as rapid and time-saving devices in the assessment of intelligence. Many of the combinations show correlations of high order with the entire test, though their use must be limited to special situations. For the most part, findings are not readily comparable for they were obtained on samples which differ considerably from each other with respect to IQ range, education, age, sex, and other factors of varying significance.

The psychologist's rejection, in recent years, of his role as a "mere psychometrist" and his aspiration to diagnose the whole personality, rather than only its intellectual component, gave rise to a strong interest in the use of the various measures as means of psychiatric diagnosis. Underlying the flood of research that was stimulated by this interest is the hypothesis that the serveral intellectual factors measured by the test are differentially affected by a variety of emotional, psychiatric, and neurological conditions. Most intratest variability seemed to have been attributed to pathological conditions without sufficient regard for such variabilities in the test profile of the everyday, nonhospitalized person. The possibility that individual differences exist in the premorbid state to such a degree that they are mistakenly thought to be amplifications due to the morbid state has not been adequately considered. Thus, group tendencies of various nosological classifications have been frequently hailed as aids in psychiatric diagnosis. In fact, the dozens of studies published have not changed much the state of affairs in respect to "scatter," described six years ago (**111**). To be sure, many more groups have been tested and compared, and some additional measures of scatter contributed. Yet the scatter mountain gave birth to a mouse. It still is true, however, that " . . . the various measures of scatter and variability—the different patterns, have succeeded in differentiating *groups*, but not *individuals*" (**111**, p. 420). The empirically devised measures still point out "trends" and "tendencies" but cannot replace clinical observations. On the clinical and "qualitative" level some of the studies have brought forth certain observations and insights that may be of practical value. There is even some evidence of greater success in diagnosis of W-B records if one is not held strictly to some particular pattern, but pools the quantitative patterns with qualitative observations. Apart from the criticism that can be leveled against the multitude of investigations for failure to study large enough groups and to control numerous important factors, the difficulty is with the underlying hypothesis itself. Implied in the hypothesis is the notion that the criterion is clear-cut and of a homogeneous nature. But, alas, not all schizophrenics are alike or obtain the same diagnosis in different hospitals; not all "organics" have the same amount and locus of brain destruction, etc., etc. The error is in assuming stability, uniformity, and homogeneity of the criterion. To do that is to build on sand.

Possibly some attention should be concentrated upon the measuring instrument itself (**46**, **75**, **78**). There are deficiencies in the structure of the test if it is to be used for the purposes just discussed. Inclusion of more significant functions and the increase in number of items within each measured function may lead to a closer approximation of the desired results. Also, experimentation with an improved scale, with subjects (preferably "normal") under varying conditions (stress, emotion, etc.) may assist in the refinement of intellectual measures which are particularly vulnerable to the more exaggerated and chronic conditions in psychopathology.

Almost any test employed by a skillful clinician and insightful worker may prove an "aid in differential diagnosis." However, so far, we have failed to demonstrate the effectiveness of certain quantitative formulas and test score variabilities as an aid in individual diagnosis. The theoretical contribution of such studies to the general field of psychology has been meager. It is doubtful that the continued clogging of the periodical literature with testimonials pro and con, based on scanty and uncontrolled evidence, will be of further benefit.

BIBLIOGRAPHY

1. Abelson, H. H., & Ellis, A. Other devices for investigating personality. *Rev. educ. Res.*, 1947, **17**, 101–109.
2. Derner, G. F., Aborn, M., & Canter, A. H. The reliability of the Wechsler-Bellevue subtests and scales. *J. consult. Psychol.*, 1950, **14**, 172–179.
3. Alexander, Frances, Crutchlow, Eileen, & Hoffman, Mary. A selective survey of the Wechsler-Bellevue section of Rapaport's *Diagnostic psychological testing. Canad. J. Psychol.*, 1947, **1**, 111–115.
4. Allen, R. M. The test performance of the brain injured. *J. clin. Psychol.*, 1947, **3**, 225–230.
5. Allen, R. M. A note on the use of the Wechsler-Bellevue Scale Mental Deterioration Index with brain injured patients. *J. clin. Psychol.*, 1948, **4**, 88–89.
6. Allen, R. M. II. The test performance of the brain diseased. *J. clin. Psychol.*, 1948, **4**, 281–284.
7. Altus, W. D. The comparative validities of two tests of general aptitude in an Army Special Training Center. *J. appl. Psychol.*, 1946, **30**, 42–44.
8. Altus, W. D. The validity of an abbreviated information test used in the Army. *J. consult. Psychol.*, 1948, **12**, 270–275.
9. Altus, W. D., & Clark, J. H. Subtest variation on the Wechsler-Bellevue for two institutionalized behavior problem groups. *J.*

consult. Psychol., 1949, **13**, 444–447.

10. Barnett, I. The use of *Z* scores in equating the Wechsler-Bellevue subtests. *J. clin. Psychol.*, 1950, **6**, 184–188.
11. Bay, Margaret, & Berks, Mary. Comprehension, similarities, and digit symbols of the Wechsler-Bellevue Scale used in a court clinic. *Amer. Psychologist*, 1948, **3**, 365. (Abstract.)
12. Blake, R., & McCarty, B. A comparative evaluation of the Bellevue-Wechsler Mental Deterioration Index distributions of Allen's brain injured patients and of normal subjects. *J. clin. Psychol.*, 1948, **4**, 415–418.
13. Boehm, Alice, & Sarason, S. Does Wechsler's formula distinguish intellectual deterioration from mental deficiency? *J. abnorm. soc. Psychol.*, 1947, **42**, 356–358.
14. Brecher, Sylvia. The value of diagnostic signs for schizophrenia on the Wechsler-Bellevue intelligence test. *Psychiat. Quart. Suppl.*, 1946, **20**, 58–64.
15. Brown, A. W. Review of Rapaport, D. *et al.*, *Diagnostic psychological testing*. *Psychol. Bull.*, 1946, **43**, 477–479.
16. Brown, M. N. A critique of the Wechsler-Bellevue system of weighted scores. *J. clin. Psychol.*, 1949, **5**, 170–173.
17. Buck, J. N. *The H-T-P* (manual). Colony, Va.: Lynchburg State Colony, 1947. (Mimeo.)
18. Carp, A. Performance on the Wechsler-Bellevue Scale and insulin shock therapy. *J. abnorm. soc. Psychol.*, 1950, **45**, 127–136.
19. Chesrow, E. J., Wasika, P. H., & Reinitz, A. H. A psychometric evaluation of aged white males. *Geriatrics*, 1949, **4**, 169–177.
20. Clark, J. H. An investigation of certain relationships between the California Test of Mental Maturity and the Wechsler-Bellevue Intelligence Scale. *J. gen. Psychol.*, 1949, **41**, 21–25.
21. Clark, J. H. Subtest variation on the Wechsler-Bellevue for two institutionalized behavior problem groups. *Amer. Psychologist*, 1949, **4**, 395. (Abstract.)
22. Cofer, C. N. Psychological test performance under hyoscine: A case of post-infectious encephalopathy. *J. gen. Psychol.*, 1947, **36**, 221–228.
23. Cohen, B. Validity of a short form of the Wechsler-Bellevue on four psychiatric groups. *Amer. Psychologist*, 1946, **1**, 464. (Abstract.)
24. Cohen, B., & Steisel, I. An evaluation of the various short forms of the Wechsler-Bellevue test. *Proc. Ia. Acad. Sci.*, 1947, **54**, 221–226.
25. Copple, G. Senescent decline on the Wechsler-Bellevue Intelligence Scale. Unpublished Doctor's dissertation, Univ. Pittsburgh, 1948.
26. Cotzin, M., & Gallagher, J. J. Validity of short forms of the Wechsler-Bellevue Scale for mental defectives. *J. consult. Psychol.*, 1949, **13**, 357–365.
27. Cummings, S. B., Jr., MacPhee, H. M., & Wright, H. F. A rapid method of estimating the IQ's of subnormal white adults. *J. Psychol.*, 1946, **21**, 81–89.
28. Cutts, R. A., & Sloan, W. Test patterns of adjusted defectives on the Wechsler-Bellevue test. *Amer. J. ment. Def.*, 1945, **50**, 98–101.
29. Cutts, R. A., & O'Kelley, Margaret. The effect of hospitalization on Wechsler-Bellevue subtest scores by mental defectives. *Amer. J. ment. Def.*, 1947, **51**, 391–393.
30. Derner, G. F., & Aborn, M. An administration board for the Wechsler-Bellevue object assembly subtest. *J. consult. Psychol.*, 1950, **14**, 71–72.

31. Diamond, S. The Wechsler-Bellevue Intelligence Scales and certain vocational aptitude tests. *J. Psychol.*, 1947, **24**, 279–282.
32. Dubin, S. S., & Thaler, M. The use of psychological tests on schizophrenic patients before and after shock treatment. *Amer. Psychologist*, 1947, **2**, 283. (Abstract.)
33. Estes, S. G. Deviations of Wechsler-Bellevue subtest scores from vocabulary level in superior adults. *J. abnorm. soc. Psychol.*, 1946, **41**, 226–228.
34. Fisher, K. A. Changes in test performance of ambulatory depressed patients undergoing electric shock therapy. *J. gen. Psychol.*, 1949, **41**, 195–232.
35. Fisher, S., & Sunukjian, Helen. Intellectual disparities in a normal group and their relationship to disturbance. *J. clin. Psychol.*, 1950, **6**, 288–290.
36. Foster, A. Age and the Wechsler-Bellevue scattergraph. *J. clin. Psychol.*, 1947, **3**, 396–397.
37. Fox, Charlotte, & Birren, J. E. Intellectual deterioration in the aged: Agreement between the Wechsler-Bellevue and the Babcock-Levy. *J. consult. Psychol.*, 1950, **14**, 305–310.
38. Franklin, J. C. Discriminative value and patterns of the Wechsler-Bellevue Scales in the examination of delinquent Negro boys. *Educ. psychol. Measmt*, 1945, **5**, 71–85.
39. Freeman, A. V. Scatter indices in schizophrenia. Unpublished Doctor's dissertation, Columbia Univ., 1948.
40. French, Elizabeth, & Hunt, W. A. The Navy Northwestern Successive Matrices Test. *Amer. Psychologist*, 1949, **4**, 268–269. (Abstract.)
41. Garfield, S. L. The Shipley-Hartford Retreat Scale as a quick measure of mental status. *J. consult. Psychol.*, 1947, **11**, 148–150.
42. Garfield, S. L. A note on Patterson's article, "The Wechsler-Bellevue Scale as an aid in psychiatric diagnosis." *J. clin. Psychol.*, 1947, **3**, 198–200.
43. Garfield, S. L. A preliminary appraisal of Wechsler-Bellevue scatter patterns in schizophrenia. *J. consult. Psychol.*, 1948, **12**, 32–36.
44. Garfield, S. L. An evaluation of Wechsler-Bellevue patterns in schizophrenia. *J. consult. Psychol.*, 1949, **13**, 279–287.
45. Garfield, S. L., & Fey, W. A comparison of the Wechsler-Bellevue and Shipley-Hartford scales as measures of mental impairment. *J. consult. Psychol.*, 1948, **12**, 259–264.
46. Garrett, H. E. Review of Rapaport, D. *et al.*, *Diagnostic psychological testing. J. abnorm. soc. Psychol.*, 1947, **42**, 134–137.
47. Geil, C. A. A clinically useful abbreviated Wechsler-Bellevue Scale. *J. Psychol.*, 1945, **20**, 101–108.
48. Gerstein, R. A. A suggested method for analyzing and extending the use of Bellevue-Wechsler vocabulary responses. *J. consult. Psychol.*, 1949, **13**, 366–370.
49. Gibby, R. G. A preliminary survey of certain aspects of Form II of the Wechsler-Bellevue Scale as compared to Form I. *J. clin. Psychol.*, 1949, **5**, 165–169.
50. Gilhooly, F. M. The relationship between variability and ability on the Wechsler-Bellevue. *J. consult. Psychol.*, 1950, **14**, 46–48.
51. Gilhooly, F. M. Correction of "The relationship between variability and ability on the Wechsler-Bellevue." *J. consult. Psychol.*, 1950, **14**, 329.
52. Gilhooly, F. M. Wechsler-Bellevue reliability and validity of certain diagnostic signs of the neuroses. *J. consult. Psychol.*, 1950, **14**, 83–87.

53. GLASER, N. M. A study of the intelligence of immigrants. *Amer. Psychologist*, 1949, **4**, 241. (Abstract.)

54. GOLDFARB, W. Adolescent performance in the Wechsler-Bellevue Intelligence Scales and the Revised Stanford-Binet examination, Form L. *J. educ. Psychol.*, 1944, **35**, 503–507.

55. GOLDMAN, R., GREENBLATT, M., & COON, G. P. Use of the Wechsler-Bellevue Scale in clinical psychiatry with particular reference to cases with brain damage. *J. nerv. ment. Dis.*, 1946, **104**, 144–179.

56. GOTHBERG, LAURA. A comparative study of the Stanford-Binet old form test and Wechsler-Bellevue, verbal performance, and full-scale, as shown in the results of unselected employees. *Amer. J. ment. Def.*, 1949, **53**, 497–503.

57. GUERTIN, W. H. Mental growth in pseudo-feeblemindedness. *J. clin. Psychol.*, 1949, **5**, 414–418.

58. GURVITZ, M. S. An alternate short form of the Wechsler-Bellevue test. *Amer. J. Orthopsychiat.*, 1945, **15**, 727–733.

59. GUTMAN, BRIGETTE. The application of the Wechsler-Bellevue Scale in the diagnosis of organic brain disorders. *J. clin. Psychol.*, 1950, **6**, 195–198.

60. HALPERN, FLORENCE. Studies of compulsive drinkers: Psychological test results. *Quart. J. Stud. Alcohol*, 1946, **6**, 468–479.

61. HAMISTER, R. The test-retest reliability of the Wechsler-Bellevue intelligence test for a neuropsychiatric population. *J. consult. Psychol.*, 1949, **13**, 39–43.

62. HARPER, E. A. Discrimination of the types of schizophrenia by the Wechsler-Bellevue Scale. *J. consult. Psychol.*, 1950, **14**, 290–296.

63. HAYS, W. A comparison of scatter patterning for mental defectives on the Wechsler Forms I and II. *Amer. J. ment. Def.*, 1950, **55**, 264–268.

64. HERMANN, M., & HACKMAN, R. B. Distribution of scales on the Wechsler-Bellevue Scales and the California Test of Mental Maturity at a VA Guidance Center. *J. appl. Psychol.*, 1948, **32**, 642–648.

65. HEWSON, LOUISE. The Wechsler-Bellevue Scale and the Substitution Test as aids in neuro-psychiatric diagnosis. *J. nerv. ment. Dis.*, 1949, **109**, 158–183; 246–265.

66. HEYER, A. W. "Scatter analysis" techniques applied to anxiety neurotics from a restricted culture-educational environment. *J. gen. Psychol.*, 1949, **40**, 155–166.

67. HODGSON, G. An analysis of subtests in the Wechsler-Bellevue verbal scale administered to 139 delinquent Mexican boys. *Amer. Psychologist*, 1948, **3**, 343. (Abstract.)

68. HOLZBERG, J. D., & DEANE, M. A. The diagnostic significance of an objective measure of intratest scatter on the Wechsler-Bellevue Intelligence Scale. *J. consult. Psychol.*, 1950, **14**, 180–188.

69. HUNT, W. A., KLEBANOFF, S. G., MENSH, I. N., & WILLIAMS, M. The validity of some abbreviated individual intelligence scales. *J. consult. Psychol.*, 1948, **12**, 48–52.

70. HUNT, W. A., FRENCH, ELIZABETH, & MENSH, I. N. An abbreviated form of the McGill Verbal Situation Test. *J. clin. Psychol.*, 1948, **4**, 406–411.

71. HUNT, W. L. The relative rates of decline of Wechsler-Bellevue "Hold" and "Don't Hold" tests. *J. consult. Psychol.*, 1949, **13**, 440–443.

72. JASTAK, J. An item analysis of the Wechsler-Bellevue tests. *J. consult. Psychol.*, 1950, **14**, 88–94.

73. JOHNSON, ANNA P. Measuring men-

tal deterioration by the differential test score method. *J. ment. Def.*, 1947, **51**, 389–390.

74. JOHNSON, L. Wechsler-Bellevue pattern analysis in schizophrenia. *J. consult. Psychol.*, 1949, **13**, 32–33.
75. JOHNSON, T. F. Some needs in research with the Wechsler-Bellevue Scale. *J. gen. Psychol.*, 1949, **41**, 33–36.
76. KASS, W. Wechsler's Mental Deterioration Index in the diagnosis of organic brain disease. *Trans. Kans. Acad. Sci.*, 1949, **52**, 66–70.
77. KESSLER, LUCILLE. Intellectual changes in schizophrenic patients following electroshock therapy. Unpublished Master's thesis, New York Univ., 1947.
78. KLEIN, G. S. An application of the multiple regression principle to clinical prediction. *J. gen. Psychol.*, 1948, **38**, 159–179.
79. KLUGMAN, S. The effect of placement of the digits test in the Wechsler-Bellevue Intelligence Scale. *J. consult. Psychol.*, 1948, **12**, 345–348.
80. KNOTT, J. R., & CANNICOTT, R. The application of brief tests of intelligence in the psychiatric clinic. *Amer. Psychologist*, 1948, **3**, 365. (Abstract.)
81. KOGAN, KATE L. The diagnosis of a patient with organic defect. *J. Personality*, 1946, **15**, 113–120.
82. KOGAN, W. S. An investigation into the relationship between psychometric patterns and psychiatric diagnosis. *J. gen. Psychol.*, 1950, **43**, 17–46.
83. KRIEGMAN, G., & HANSEN, F. W. VIBS: A short form of the Wechsler-Bellevue Intelligence Scale. *J. clin. Psychol.*, 1947, **3**, 209–216.
84. KRUGMAN, M. Review of Rapaport, D., *et al.*, *Diagnostic psychological testing*. *J. consult. Psychol.*, 1946, **10**, 285–286.
85. KUTASH, S. B. A comparison of the Wechsler-Bellevue and Revised Stanford-Binet scale for adult defective delinquents. *Psychiat. Quart.*, 1945, **19**, 677–685.
86. LEVI, J., OPPENHEIM, S., & WECHSLER, D. Clinical use of the Mental Deterioration Index of the Bellevue-Wechsler Scale. *J. abnorm. soc. Psychol.*, 1945, **40**, 405–407.
87. LEVINE, J., & BLACKBURN, A. R. Intelligence test scores of newly blinded soldiers. *J. consult. Psychol.*, 1950, **14**, 311–315.
88. LEVINE, L. S. The utility of Wechsler's pattern in the diagnosis of schizophrenia. *J. consult. Psychol.*, 1949, **13**, 28–31.
89. LEWINSKI, R. J. Discriminative value of the sub-tests of the Bellevue verbal scale in the examination of naval recruits. *J. gen. Psychol.*, 1944, **31**, 95–99.
90. LEWINSKI, R. J. The psychometric pattern: I. Anxiety neurosis. *J. clin. Psychol.*, 1945, **1**, 214–221.
91. LEWINSKI, R. J. The psychometric pattern: II. Migraine. *Psychiat. Quart.*, 1945, **19**, 368–376.
92. LEWINSKI, R. J. The psychometric pattern: III. Epilepsy. *Amer. J. Orthopsychiat.*, 1947, **17**, 714–722.
93. LEWINSKI, R. J. Vocabulary and mental measurement: A quantitative investigation and review of research. *J. genet. Psychol.*, 1948, **72**, 247–281.
94. LINDNER, R. M., & GURVITZ, M. Restandardization of the Revised Beta Examination to yield the Wechsler type of I.Q. *J. appl. Psychol.*, 1946, **30**, 649–658.
95. LORGE, I. Intellectual changes during maturity and old age. *Rev. educ. Res.*, 1947, **17**, 326–332.
96. LUBORSKY, L. Psychometric changes during electric shock treatment. *J. nerv. ment. Dis.*, 1948, **107**, 531–536.

97. Luchins, A. S. On certain misuses of the Wechsler-Bellevue Scales. *J. consult. Psychol.*, 1946, **10**, 109–111.

98. Madonick, M. J., & Solomon, M. The Wechsler-Bellevue Scale in individuals past sixty. *Geriatrics*, 1947, **2**, 34–40.

99. Magaret, Ann, & Simpson, Mary. A comparison of two measures of deterioration in psychotics. *J. consult. Psychol.*, 1948, **12**, 265–270.

100. Mayman, M. An analysis of scatter in intelligence test results: A review of the literature. *Trans. Kans. Acad. Sci.*, 1946, **48**, 429–444.

101. McCullough, M. W. Wechsler-Bellevue changes following prefrontal lobotomy. *J. clin. Psychol.*, 1950, **6**, 270–273.

102. McNemar, Q. On abbreviated Wechsler-Bellevue scales. *J. consult. Psychol.*, 1950, **14**, 79–81.

103. McNemar, Q. Review of Rapaport, D., *et al.*, *Diagnostic psychological testing*. *Amer. J. Psychol.*, 1946, **59**, 306–311.

104. MacPhee, H. M., Wright, H. F., & Cummings, S. B., Jr. The performance of mentally subnormal rural southern Negroes on the verbal scale of the Bellevue intelligence examination. *J. soc. Psychol.*, 1947, **25**, 217–229.

105. Olch, Doris. Psychometric pattern of schizophrenics on the Wechsler-Bellevue intelligence test. *J. consult. Psychol.*, 1948, **12**, 127–136.

106. Pascal, G., & Zeaman, Jean. A note on the validity of Wechsler-Bellevue scatter. *Amer. J. Psychiat.*, 1949, **105**, 840–842.

107. Patterson, C. H. A comparison of various "short forms" of the Wechsler-Bellevue Scale. *J. consult. Psychol.*, 1946, **10**, 260–267.

108. Patterson, C. H. The Wechsler-Bellevue Scale as an aid in psychiatric diagnosis. *J. clin. Psychol.*, 1946, **2**, 348–353.

109. Patterson, C. H. A further study of two short forms of the Wechsler-Bellevue Scale. *J. consult. Psychol.*, 1948, **12**, 147–152.

110. Peixotto, Helen. Wechsler-Bellevue patterns: A note of caution. *J. clin. Psychol.*, 1950, **6**, 188–190.

111. Rabin, A. I. The use of the Wechsler-Bellevue Scales with normal and abnormal persons. *Psychol. Bull.*, 1945, **42**, 410–422.

112. Rabin, A. I., Davis, J. C., & Sanderson, M. H. Item difficulty on some Wechsler-Bellevue subtests. *J. appl. Psychol.*, 1946, **30**, 493–500.

113. Rabin, A. I. Effects of electroshock treatment upon some aspects of personality and intellect. *Amer. Psychologist*, 1947, **2**, 204. (Abstract.)

114. Rapaport, D., Gill, M., & Schafer, R. *Diagnostic psychological testing*, Vol. I. Chicago: Year Book Publishers, 1945.

115. Rappaport, S. R., & Webb, W. B. An attempt to study intellectual deterioration by premorbid and psychotic testing. *J. consult. Psychol.*, 1950, **14**, 95–98.

116. Rashkis, H. A., & Welsh, G. S. Detection of anxiety by use of the Wechsler Scale. *J. clin. Psychol.*, 1946, **2**, 354–357.

117. Rogers, L. S. A comparative evaluation of the Wechsler-Bellevue Mental Deterioration Index for various adult groups. *J. clin. Psychol.*, 1950, **6**, 199–202.

118. Rogers, L. S. A note on Allen's index of deterioration. *J. clin. Psychol.*, 1950, **6**, 203.

119. Rubinstein, E. A note on recording block design performance of the Wechsler-Bellevue Scales. *J. clin. Psychol.*, 1948, **4**, 307–308.

120. Sappenfield, B. R. Rapid method for placement of Wechsler object

assembly pieces. *J. clin. Psychol.*, 1947, **3**, 301.

121. Sartain, A. Q. A comparison of the New Revised Stanford-Binet, the Bellevue Scale, and certain group tests of intelligence. *J. soc. Psychol.*, 1946, **23**, 237–239.
122. Schlosser, J. R., & Kantor, R. E. Wechsler's deterioration ratio in psychoneurosis and schizophrenia. *J. consult. Psychol.*, 1949, **13**, 108–110.
123. Schmidt, D. G. Levels of intelligence of prison inmates. *Amer. J. ment. Def.*, 1946, **51**, 63–66.
124. Shoben, E. J. The Wechsler-Bellevue in the detection of anxiety: A test of the Rashkis-Welsh hypothesis. *J. consult. Psychol.*, 1950, **14**, 40–45.
125. Sloan, W. Validity of Wechsler's deterioration quotient in high grade mental defectives. *J. clin. Psychol.*, 1947, **3**, 287–288.
126. Sloan, W., & Cutts, R. A. Test patterns of defective delinquents on the Wechsler-Bellevue test. *Amer. J. ment. Def.*, 1945, **50**, 95–97.
127. Sloan, W., & Guertin, W. H. A comparison of H-T-P and Wechsler-Bellevue IQ's in mental defectives. *J. clin. Psychol.*, 1948, **4**, 424–426.
128. Sorsby, F. A briefer method for the Wechsler-Bellevue vocabulary subtest. Unpublished Master's thesis, Univ. Pittsburgh, 1949.
129. Spache, G. Scoring qualitative responses on the Wechsler-Bellevue Scale. *Amer. J. Orthopsychiat.*, 1948, **18**, 360–363.
130. Springer, N. N. A short form of the Wechsler-Bellevue intelligence test as applied to Naval personnel. *Amer. J. Orthopsychiat.*, 1946, **16**, 341–344.
131. Stacey, C. L., & Levin, Janice. Performance of retarded individuals on Stanford-Binet and Wechsler-Bellevue intelligence scales. *Amer. J. ment. Def.*, 1950, **55**, 123–132.
132. Strother, C. R. The performance of psychopaths on the Wechsler-Bellevue test. *Proc. Ia. Acad. Sci.*, 1944, **51**, 397–400.
133. Tagiuri, R. Comparison of results obtained from the Wechsler-Bellevue vocabulary test with those from the Stanford-Binet vocabulary test using a population of normal subjects and mental patients. *Bull. Canad. Psychol. Ass.*, 1946, **6**, 101. (Abstract.)
134. Teicher, M. I., & Singer, E. A report on the use of the Wechsler-Bellevue Scales in an overseas general hospital. *Amer. J. Psychiat.*, 1946, **103**, 91–93.
135. Warner, S. J. The Wechsler-Bellevue psychometric pattern in anxiety neurosis. *J. consult. Psychol.*, 1950, **14**, 297–304.
136. Watson, R. I. The use of the Wechsler-Bellevue Scales: A supplement. *Psychol. Bull.*, 1946, **43**, 61–68.
137. Webb, W. B. A note on the Rabin ratio. *J. consult. Psychol.*, 1947, **11**, 107–108.
138. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams and Wilkins, 1944.
139. Welsh, G. S. A note on scoring Wechsler-Bellevue subtests. *J. clin. Psychol.*, 1949, **5**, 421–422.
140. Whiteman, M. Altitude as a reference point in scatter analysis. *J. clin. Psychol.*, 1950, **6**, 160–164.
141. Williams, M., Hunt, W. A., & French, E. A further standardization of the CVS abbreviated intelligence scale. *Amer. Psychologist*, 1948, **3**, 365. (Abstract.)
142. Wittenborn, J. R. An evaluation of the use of Bellevue-Wechsler subtest scores as an aid in psychiatric

diagnosis. *J. consult. Psychol.*, 1949, **13**, 433–439.

143. Wright, H. F., MacPhee, H. M., & Cummings, S. B., Jr. The relationship between the Kent EGY and the Bellevue verbal scales. *J. abnorm. soc. Psychol.*, 1949, **44**, 223–230.

144. Wright, M. E. Use of the Shipley-Hartford Test in evaluating intellectual functioning of neuropsychiatric patients. *J. appl. Psychol.*, 1946, **30**, 45–50.

145. Yarcozynski, G., Boshes, B., & Davis, L. Psychological changes produced by frontal lobotomy. In J. F. Fulton *et al.*, *The frontal lobes.* Baltimore: Williams and Wilkins, 1948.

Received July 28, 1950.

A PSYCHOPHYSICAL METHOD EMPLOYING A MODIFICATION OF THE MÜLLER-URBAN WEIGHTS[1]

STEPHEN HARRISON
General Foods Corporation
AND
MARGARET J. HARRISON

The proposals made in this short paper have arisen out of taste and flavor threshold determinations carried out at these laboratories, but they should be applicable to psychophysics in general.

We have recently had the task of determining stimulus (absolute) thresholds for a number of different substances. Sometimes the threshold was measured for addition to spring water, but on some occasions it was necessary for our purposes to utilize an infusion of coffee as a base.

It has been the experience of this and many other laboratories working in the same field that organoleptic judgments are particularly prone to the "stimulus error"—i.e., the taster tends to align his judgment to his knowledge of the stimulus (2). It was, therefore, taken as an essential prerequisite of any procedure adopted that all samples should be coded before presentation to the subject. This restriction somewhat limited the choice of psychophysical methods, since it is a feature of much of the classical psychophysical procedure for absolute threshold determination that the subject has either knowledge or part knowledge of each stimulus presented to him. Stratton had proposed the method of serial groups, in which each stimulus value was presented together with an equal number of blank or "catch" stimuli, the task of the subject being to recognize the true stimuli. The mathematical treatment of the results leaves much to be desired, however, as the data are not fully utilized in deriving the threshold value and no parameter equivalent to a standard error is provided (7).

We have, therefore, developed the following method of estimation. In essence, coded beakers are presented in a series of pairs to the subject, each pair consisting of one beaker containing one of five concentration levels of the active substance under investigation and a second beaker serving as a "blank" stimulus containing water or other medium upon which the threshold determination is being based. No attempt will be made here to describe the details of the procedure, which is in accordance with accepted psychophysical practice and follows recognized experimental design. The data are collected in the form of percentage of correct responses for each stimulus level presented. The mathematical treatment of the data, however, involves certain innovations in advance of standard methods which it is the purpose of this paper to present.

---

[1] The writers wish to acknowledge the cooperation of Dr. L. W. Elder, Director of the Produce Evaluation Laboratory of this organization, and of his test kitchen staff supervised by Mrs. A. M. Chao.

METHOD OF DETERMINING THE THRESHOLD

The method consists essentially in fitting a line to the data (after adjustment as described below), assuming the phi-log-gamma hypothesis (which is preferable to the phi-gamma hypothesis owing to the large value usually found for Weber's fraction for olfactory and gustatory modalities [8]). The technique of line-fitting used is that of least squares, as described, for example, by Guilford (6), but with three modifications.

1. A feature of the probit method as described by Bliss (3) and referred to by Finney (5) is incorporated into the analysis. This employs the least squares solution based upon an eye-fitted line. This has the particular advantage of permitting the use of zero and 100 per cent correct responses, which is not possible in the ordinary least squares treatment as described in the psychophysics literature.

2. A special allowance has to be made for the fact that a score of 50 per cent correct judgments would be expected by chance alone for each stimulus level. The original score should, therefore, be given a simple transformation so that it may express the percentage of the maximum possible score after chance expectancy has been discounted. This may be done as follows:

$$P = 2(p - 50)\%,$$

where P is the transformed percentage and p the original percentage of correct responses. The transformed percentages will then range from zero to 100 per cent (where zero indicates complete absence of discrimination and 100 per cent indicates perfect discrimination) provided the convention is adopted of interpreting negative transformed percentages as zero. This appears to the writers to be a reasonable convention since there would seem to be no way in which a taster could have negative discrimination. That is to say, it is assumed that original scores of less than 50 per cent (giving negative transformed percentages) can arise only from chance variations.

3. The Müller-Urban weights as normally employed in the constant process will clearly not be applicable here. The reason for this is as follows:

The Müller weight has the function of compensating for the fact that fluctuations of scores in the region of the threshold (i.e., the 50 per cent point) have less of a disturbing effect upon the location of the threshold value than high or low scores (4). Müller suggested that this should be compensated for by means of a weighting system based upon the squared ordinate of the normal curve. Hence, the Müller weights can properly be applied to the transformed percentage scores here proposed, since the transformed percentage scale ranges from zero to 100 per cent with the threshold at 50 per cent (whereas the scale before transformation ranges from 50 per cent—effectively—to 100 per cent, with the threshold point at 75 per cent). The Urban weights correct for the fact that the error variance of a percentage score is not a constant but varies with the magnitude of the score, being greatest at the 50 per cent point and diminishing to zero as the score approaches zero or 100 per cent (4). Urban suggested that this should be compensated for by basing a weighting system on the reciprocal of the variance of a percentage, i.e., $1/pq$. This latter system of weights clearly cannot be applied *directly* to the transformed percentages since the error variances upon which the Urban weights are based depend upon the *original* rather than the transformed percentages. Normally, the Müller and Urban weights are multiplied together to give a combined Müller-Urban weight, and it is in this form that they are usually tabled for use. These will not be applicable to the data under consideration since, as explained above, the Müller weight should properly be applied to the transformed percentage scores and the Urban

TABLE I
MODIFIED MÜLLER-URBAN WEIGHTS

P %	*Weight*	*P* %	*Weight*	*P* %	*Weight*	*P* %	*Weight*
1	0.003	26	0.475	51	0.904	76	0.962
2	.010	27	.495	52	.915	77	.952
3	.019	28	.512	53	.925	78	.941
4	.031	29	.538	54	.935	79	.929
5	.045	30	.559	55	.944	80	.915
6	.060	31	.579	56	.953	81	.903
7	.076	32	.599	57	.960	82	.884
8	.094	33	.619	58	.968	83	.866
9	.112	34	.638	59	.974	84	.845
10	.131	35	.657	60	.980	85	.824
11	.151	36	.675	61	.986	86	.800
12	.171	37	.694	62	.990	87	.773
13	.191	38	.712	63	.994	88	.745
14	.213	39	.730	64	.996	89	.715
15	.234	40	.747	65	0.999	90	.681
16	.256	41	.763	66	1.000	91	.645
17	.277	42	.780	67	1.000	92	.605
18	.299	43	.796	68	1.000	93	.561
19	.321	44	.811	69	0.999	94	.512
20	.343	45	.826	70	.997	95	.458
21	.365	46	.840	71	.993	96	.399
22	.387	47	.854	72	.989	97	.329
23	.409	48	.867	73	.984	98	.249
24	.431	49	.880	74	.978	99	.151
25	0.453	50	0.892	75	0.970	100	0.000

$P = 2(p - 50)\%$, where p is the original percentage of correct responses.

weights to the original percentage scores. Instead, therefore, of using the normal Müller-Urban weights, it is suggested that a modified product of the two weights be used where the Müller weight is based upon the *transformed* percentage score and the Urban weight on the *original* percentage score. A table of these weights has been constructed for scores ranging from zero to 100 per cent at 1 per cent intervals (see Table I). To make these weights directly comparable to the ordinary Müller-Urban weights, they have beeen adjusted so as to bring the maximum weight up to unity.

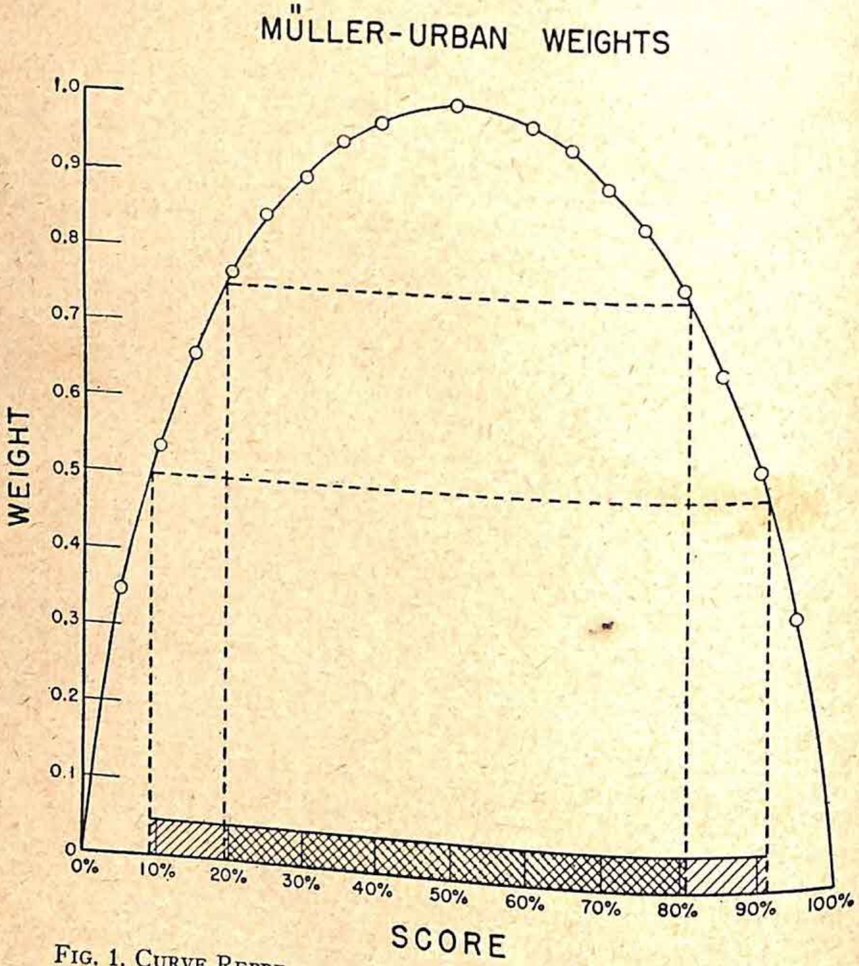

FIG. 1. CURVE REPRESENTING STANDARD MÜLLER-URBAN WEIGHTS

The standard procedure for the calculation of the threshold as given by Guilford (6) or by Bliss (3) has been followed except that calculations are based upon transformed percentages, and the modified Müller-Urban weights are used instead of the standard weights. These calculations provide a measure of the standard error as well as the threshold, which makes it possible to establish fiducial limits for the threshold. This seems to the present writers to be a matter of great importance since the estimate of the threshold has little value by itself owing to the large

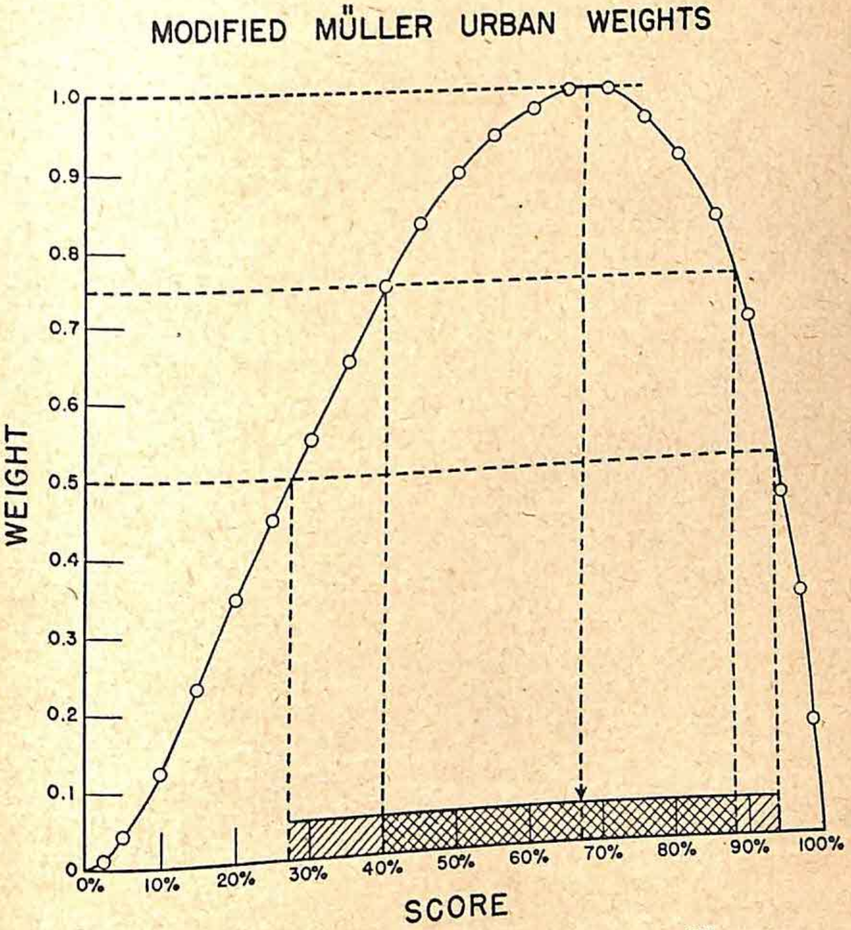

FIG. 2. CURVE REPRESENTING MODIFIED MÜLLER-URBAN WEIGHTS
The percentage score has been given transformation.

standard errors characteristic of thresholds in the fields of taste and flavor. A threshold determination is of little practical use unless it can be contained within fiducial limits. It has been chiefly this circumstance which has led the authors to develop this modified psychophysical procedure.

COMPARISON OF THE NEW WEIGHTS WITH THE NORMAL MÜLLER-URBAN WEIGHTS

It is interesting to compare the properties of this new weighting distribution with the normal Müller-Urban weights (see Figures 1 and 2). The following differences may be noted:

1. The maximum weight is no longer found at the threshold (50 per cent) point, but is displaced positively, i.e., towards the upper end of the scale, falling at about 67 per cent.

2. The weights corresponding to the lower half of the distribution are much smaller than in the case of the standard Müller-Urban weights, owing to the high error variances of the corresponding original scores in this region.

3. The spread of scores having high weights is less than in the case of the standard Müller-Urban values. This is shown on the two figures. The heavily shaded area in each case includes all scores carrying a weight at least three-quarters of the maximum value. The spread is only 40%–87% = 47% for the modified weights, compared with 19%–81% = 62% for the standard weights. Similarly, the half-shaded area in each case includes all scores carrying a weight of at least half the maximum value. For the modified weights, the spread is only 27%–94% = 67%, compared with 9%–91% = 82% for the standard weights. Hence, it would appear desirable for the transformed scores to fall within the range 27%–94% if possible in order that each score may receive a high weighting factor. However, too restricted a range has been found undesirable since this gives a poor reliability for the slope of the regression line and hence, particularly, for the standard error of concentrations corresponding to the more extreme response scores.

The restricted range of scores of high weight for the modified weights can be expanded if the chance expectation of correct responses is reduced below the 50 per cent level by any suitable means. The argument may be stated in general terms as follows. Where p represents the original percentage of correct responses and e the percentage of chance expectation, then the percentage score after allowance for the chance expectation will be given by

$$P = (p - e)\left\{\frac{100}{100 - e}\right\}$$

so that

$$p = \frac{P(100 - e)}{100} + e.$$

The proposed weight is given by

$$\frac{y^2}{pq},$$

where y is the ordinate of the Gaussian curve of errors, the value of y being determined by the transformed percentage P; $q = 100 - p$.

Expressing the denominator in terms of the transformed percentage P, the expression becomes

$$\frac{y^2}{\left\{\frac{P(100 - e)}{100} + e\right\}\left\{100 - \frac{P(100 - e)}{100} - e\right\}}$$

which may be written in the form

$$\frac{y^2}{\left(\frac{100 - e}{100}\right)^2\left\{P + \frac{100e}{(100 - e)}\right\}(100 - P)}.$$

The squared expression in the denominator will be the same for all weights (for a given value of e) and so can be eliminated without affecting the relative values of the weights. The formula for the relative weights is, therefore, given by

$$\frac{y^2}{\left\{P + \frac{100e}{(100 - e)}\right\}(100 - P)}.$$

The numerator and denominator components of this weight now apply uniformly to the transformed percentage values.

As one means of reducing the chance expectation, the so-called triangular procedure might be utilized in which samples are presented in groups of three, two of which contain identical concentrations of the stimulus substance and the third, a blank solution only, the task of the subject being to identify the odd sample (1). Chance expectancy would then fall to $33\frac{1}{3}$ per cent, but the subject would have to examine one and one-half times as many samples as in the case of the method described above. In the field of organoleptics, onset of fatigue is particularly rapid, and for this reason we have not so far used the triangular test in systematic threshold determinations.

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated
Acc.

BIBLIOGRAPHY

1. Bengtsson, K. Innebörden i statischt utslag vid provemakning. (Statistical evaluation of taste test results.) *Svensk bryggeritidskrift*, 1946, No. 7.
2. Bengtsson, K., & Helm, E. Principles of taste testing. *Wallerstein Communications*, 1946, 9, 171–180.
3. Bliss, C. I. The calculation of the dosage-mortality curve. *Ann. appl. Biol.*, 1935, 22, 134–167.
4. Boring, E. G. Urban's tables and the method of constant stimuli. *Amer. J. Psychol.*, 1917, 28, 280–293.
5. Finney, D. J. The application of probit analysis to the results of mental tests. *Psychometrika*, 1944, 9, 31–39.
6. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
7. Stratton, G. M. The method of serial groups. *Psychol. Rev.*, 1902, 9, 444–447.
8. Thurstone, L. L. The phi-gamma hypothesis. *J. exp. Psychol.*, 1928, 11, 293–305.

Received July 6, 1950.

SOME COMMENTS ON THE CONTROLLING OF DIFFERENCES IN RESPONSES ON THE RORSCHACH TEST[1]

ROBERT M. EICHLER[2]
State University of Iowa

In a recent article in this journal, Cronbach (1) has called attention to the need for controlling individual differences in the number of responses given to the Rorschach test before proceeding to test the significance of relations involving the remaining Rorschach scores. As he indicates, the experimental dependence of the various Rorschach scores on "productivity" or number of responses (R) may result in a repetition of the test of the significance of relations involving differences in R when functions concerning other Rorschach categories are evaluated.

Since the typical method for controlling productivity differences by converting Rorschach scores to percentages presents serious statistical complications and, in addition, often fails to achieve the desired independence from R, Cronbach proposes several alternative procedures for resolving this difficulty. These are:

1. Scoring only a fixed number of responses on all protocols.
2. Reconstructing subgroups equated on R either by discarding cases or establishing subsamples for various intervals in which R is nearly uniform.
3. Plotting the variable against R and testing the significance of the proportion of cases falling above and below a line fitting the medians of the columns.
4. A method of "pattern tabulation" involving the analysis of profiles of normalized scores.

Cronbach's discussion of this particular problem and its possible solutions suggests two additional points that need consideration. The first relates to the inclusion of Rorschach scores in R (the problem of part-whole correlation); the second concerns the use of analysis of covariance with transformed Rorschach scores.

The dependence of a Rorschach score upon total responsiveness has usually been shown by correlating R and that score. Since the score is reflected in R, however, it is to be expected that the two will be positively correlated. It may be wondered, therefore, whether the Rorschach score is not actually independent of productivity when the latter does not include it. If this is true, the researcher would not be confronted with the task of controlling for R differences, before systematically testing his hypotheses concerning the other test categories. Evidence obtained by the writer (2) indicates that, even with total productivity freed of a particular score, the correlation between the two tends to be quite high. For example, the study (60 cases) revealed correlations between weighted shading responses and R *minus* shading responses of .53, between F and R *minus* F of .49, between weighted color responses and R *minus* color responses of .38, between P and R *minus* P of .37. Thus, even when one avoids any spurious correlation, a "fluency" factor is apparent. In view of these significant relationships it is manifestly an important first step, as Cronbach suggests, to control for productivity differences if obtained differences in other Rorschach categories are to be interpreted soundly.

---

[1] The writer is indebted to Dr. Harold P. Bechtoldt for his helpful suggestions.

[2] Public Health Service Postdoctorate Research Fellow of the National Institute of Mental Health. Now at the New York Regional Office of the Veterans Administration.

In his discussion of procedures for achieving independence from R, Cronbach has apparently not considered the possibility of using analysis of covariance. The techniques he suggests require either a large number of cases or a considerable number of responses per blot if they are to be useful. Since the administration and scoring of a Rorschach test demand a good deal of time, the usual researcher in this area generally uses a sample fairly limited in size, with the resultant N for subgroups even more limited. For such investigations covariance analysis would seem to be better adapted to handle the problem of productivity control. Such analysis makes the groups comparable on the productivity factor by determining how much of the variance in the particular Rorschach score can be predicted from total responsiveness, and then subtracting this to secure the residual variance as the adjusted value. It should be noted that if this method of analysis is employed, it will be generally necessary to normalize the data by some appropriate transformation, as most Rorschach scores are considerably skewed. The writer has found the square root transformation (4) a useful method in this regard. To avoid spurious correlations between productivity and other Rorschach scores, it is also suggested that the total number of responses *minus* the number of responses in the dependent variable (the Rorschach score under analysis) be used as the independent or uncontrolled variable.

Some qualifications concerning the use of analysis of covariance should be made. In a simple covariance design where the total variance is analyzed into the methods and within groups components, the extent to which this technique will increase the precision of an experiment will depend upon the within groups correlation between productivity and the Rorschach score. Where this correlation is low and/or where the group differences on the unadjusted Rorschach score under analysis are small and attributable to chance, the use of this design may be of little utility. In cases where a more complex covariance design is used involving the interaction variance as the error term, the increase in precision obtained will depend upon the correlation of the interaction effects (3). Such a design will not be profitable where the latter correlation is low.

BIBLIOGRAPHY

1. Cronbach, L. J. Statistical methods applied to Rorschach scores: A review. *Psychol. Bull.*, 1949, 46, 393–431.
2. Eichler, R. M. Experimental stress and alleged Rorschach indices of anxiety. *J. abnorm. soc. Psychol.*, 1951, 46. (In press.)
3. Lindquist, E. F. *Statistical analysis in educational research.* Boston: Houghton Mifflin, 1940.
4. Snedecor, G. W. *Statistical methods.* Ames, Iowa: Iowa State Coll. Press, 1946.

Received July 5, 1950.

Boring, E. G. *A history of experimental psychology.* (2nd Ed.) New York: Appleton-Century-Crofts, 1950. Pp. xxi+777. $6.00.

The second edition of his *History* is, in Boring's own words, "about one-third larger than the first edition, is one-half new writing and uses for its other half about two-thirds of the old edition. There must be some thousand minor changes in what is reprinted. . . . [Despite the changes] the old rule for the first edition and for my *Sensation and Perception in the History of Experimental Psychology* still holds: I speak with confidence up to twenty years ago (1930); I speak, but with less assurance, of the next decade; whatever I say for the most recent decade is based on gratuitous courage."

Changes have indeed been made. Chapter 11, in the early part of the book, is new and presents material on the Scottish school, French empiricism, and French materialism. The material of chapter 13 is extended by a consideration of Kant; and Hering receives amplified treatment in chapter 17.

The greatest changes occur in the last part. Chapters 22 through 26, dealing, respectively, with American functional psychology, Gestalt psychology, behavioristics, brain function, and dynamic psychology, have been completely rewritten.

Chapter 27, entitled "Retrospect," gives a valuable summary of important men and trends in the history of psychology. In assessing the great contributions, Boring says: "The student of the history of psychology would not be promoting an absurdity if he placed on the horizon of his imagination these three landmarks: Fechner's *Elemente der Psychophysik* of 1860, Ebbinghaus' *Über das Gedächtnis* of 1885, and Freud's *Die Traumdeutung* of 1900." (Many psychologists will want to know why Pavlov's *Conditioned Reflexes* is not included.)

In his 1929 edition Boring failed to find any great psychologists; now he says that there were four: Darwin, Helmholtz, James, and Freud. [I hereby nominate Pavlov!] In 1950, however, the *great-man* theory of scientific progress is abandoned; the great man is now regarded as only one of the agents of progress.

Although he states that he cannot speak confidently of the years since 1930, Boring gives us (in his Preface and in individual chapters, but especially chap. 27) an account of psychology in 1950. These are the striking elements: All American psychology is functional. Gestalt psychology has accomplished its mission and is now gone. Behaviorism and operationism have placed a stamp of objectivity upon American psychology. [It is disappointing not to find an account of H. M. Johnson's pioneer operational analyses (*Psychol. Rev.*, 1928, 1929, 1930).] Studies of brain function still go on [but little is said of their present significance]. Dynamic psychology is an interest, a concern, and a movement for "those psychologists who, disappointed in the description of consciousness and behavior, seek a more satisfying psychology of what they sometimes call 'human nature' "; it appeals to many psychologists; hence it constitutes a strong movement.

In 1950, psychology need make no apology. Through the years psychology has gained self-confidence from the successful application of its facts and principles. All areas of psychology show encouraging growth. "In both Great Britain and America psychology's battle has been to become independent of philosophy. . . . That conflict between parent fixation and the need for independence was still felt in 1929, but now it has all gone." [But think of the number of courses in scientific method that are given in departments of psychology! Think also of the annual crop of "systematic" articles!]

One of the first things that impresses a reviewer in comparing the first and second editions of the *History* is the similarity of historical method used in the two volumes. The *great-man* theory is disavowed in 1950, but the historical analysis exemplified in the second edition is not greatly changed from that of the 1929 edition, where, presumably, the theory helped to determine the treatment.

It may have been the *great-man* view that produced, in the 1929 edition, a history emphasizing men, their controversies, and their secondary professional activities. The weighting of emphasis led Woodworth (*J. genet. Psychol.*, 1930, 38, 521–526), in his review of the first edition, to say:

> The author's story of the past seventy years centers about the systematic problems and difficulties. He has much to say of the busy externals of psychology, the founding of laboratories, journals, and societies in the different countries, and the succession of important systematic books that have appeared. He is particularly full on the biographical side. To quite an extent, he presents the broadening of the scope of psychology as a consequence of the new methods invented. But he says little of results. He does not show how the psychologists increased the body of knowledge inherited from the sense physiologists, nor what new knowledge they have gained with their own new methods.

The situation is not greatly changed in the second edition. The *great-man* theory, although disavowed, has left its imprint.

Fortunately, we do not now have to share in Woodworth's frustration of 1930. Other works (for example, Murphy's *Historical Introduction to Modern Psychology*, are available to fill out the historical picture, and, in particular, we have Boring's *Sensation and Perception in the History of Experimental Psychology*, an account that treats experimental psychology as a factual science. The *History* and *Sensation and Perception* are complementary; and if a third book on learning and motivation should one day make its appearance (a possibility that Boring mentioned in his Preface to *Sensation and Perception*) we shall have a trilogy that gives an admirable account of a large segment of experimental psychology, written in a temporal context that varies, depending on the subject matter, from 1930 to the present.

Boring's insistence that he cannot speak confidently of the years since 1930 poses a difficulty for a reviewer: Should the reviewer ignore all that Boring says concerning the last twenty years, or should he give it serious attention? The appropriate answer seems to be that the discussion is on the record; hence it cannot be ignored. A more indirect answer argues for its consideration because it provides a clue to the vantage point from which history was surveyed; hence it presents a background against which Boring's historical selections and emphases may be evaluated.

A history written in the context of 1930 may be expected to be different from one written in a context of 1950. Further, it may be expected that if the historian shifts his vantage point from time to time and from topic to topic within the twenty-year span, the shifting point of reference will give rise to different interpretations of historical trends.

Boring's tendency to shift his point of survey within the twenty-year span since 1930 is sometimes disconcerting. It is true that, in some instances, a more sympathetic account of a topic is achieved by considering its history in the context of 1930. (For example, it is likely that the contributions of the introspective psychologists will be regarded much more sympathetically in the earlier context than in the later. It is hard to imagine that the controversies which concerned these men hold great interest for us now, nor are we likely to get excited about the philosophical backgrounds of many of the controversies.) But in other instances, as, for example, the discussion of vision in both the *History* and *Sensation and Perception*, one would like a discussion of more recent material. (Hecht, whose work far outweighs in importance the contributions of most of the earlier workers, is not mentioned.) Without the discussion of recent results, the significance of the account is diminished. In any case, the temporal unevenness of the topical casting of accounts leaves us with a feeling that a considerable reworking of history since 1930 will be required before we have a rounded picture of the events of the last two decades and, possibly more important, before we can begin to visualize the effects of present points of reference upon the interpretation of earlier psychology.

However, a few generalizations seem to be in order. The years near 1930 were important years in the history of psychology. Psychology at that time was beginning to feel some influences that had remained small or latent, and the decade ending with 1930 marked the rapid recession from the high water mark of introspective psychology. Experimental psychology before 1930 was introspective; since 1930 it has been objective.

In establishing the objective nature of psychology, one influence is of great importance, the biological. Boring does not give a rounded account of this influence. Names are listed: Helmholtz, Pavlov, Loeb, Sherrington, Jennings, Adrian, and those of many other men; but the influences of the men named are not as clearly traced as are the influences of the psychologists. Psychologists since 1930 may deal in molar concepts, but in one main regard and in many minor regards, their approach is biological: they deal with behavior and not with mind. In the study of behavior, many basic concepts and the emphasis on biological, objective method come from our biological rather than our psychological ancestors. Even the work on so-called sensation and perception, which might be expected to trace most readily back to psychological influences, is probably more influenced by the work of the sensory physiologists than it is by the work of sensory psychologists. And so in all areas of experimental psychology a primary legacy of methodology and concept is derived from a line of biological workers, many of them outside the general stream of early psychological history. In some future history it may be as necessary to trace those biological trends that helped determine American objective psychology as it is to follow the long path of introspective psychology. It may turn out that this type of history will emphasize such things as the way experiments were done and the way questions were asked rather than verbalized tenets of faith.

Another important influence that may be traced in a later history is the influence of psychiatry and the social sciences on experimental psychology. This influence may, in this volume, be pieced together on the basis of the discussion of many individuals and the chapter on dynamic psychology, but many psychologists will probably feel that nothing short of an account fully commensurate with that given introspective psychology will suffice. There can be little doubt that medicine and social science have had some influence on experimental psychology, and insofar as the general body of psychologists is concerned, one needs only to look at interests represented in the American Psychological Association to be convinced that, as of 1950, G. Stanley Hall has carried the day.

One cannot but be impressed anew with Boring's scholarship. It is probably true that no other psychologist could provide the high level of historical narration that Boring demonstrates in his *History*. The presentation is lively, the attention to detail meticulous, and the array of facts overwhelming. As Elliott says in the editor's introduction: "Even if psychologists are a crotchety lot, no audacious critic among them is going to rise and deny that Boring's *History* is a classic." The *History* is indeed a classic; but its outstanding quality derives more from an unparalleled treatment of the rise and fall of introspective psychology than it does from a delineation of the roots of objective psychology.

C. H. GRAHAM.

Columbia University.

ENGLE, T. L. *Psychology: Its principles and applications.* (Rev. Ed.) New York: World Book Co., 1950. Pp. xi+628. $3.08.

Authors of texts in elementary psychology have generally ignored the increasing interest in this subject as part of the high school curriculum. All too frequently general educational values inherent in the study of general psychology have been sacrificed in an attempt to provide an introduction to knowledge and skills thought to be the necessary equipment of the bright college sophomore who might happily decide to complete an undergraduate major and eventually to acquire the Ph.D. degree in psychology.

This book differs from the usual introductory text in that it has been written for general students in both the final years of high school and the beginning years of college, who in most instances will have no further courses in the subject. It is a revision of the first edition (1945) and incorporates numerous modifications based on suggestions of teachers, questions and comments raised by the author's students, and the results of psychological research performed during and following World War II.

The 17 chapters are organized into six major units: "I. What is the Science of Psychology," "II. Patterns of Human Behavior," "III. Why We Have Our Varied Patterns of Behavior," "IV. Learning," "V. Problems of Personal Adjustment," "VI. Broader Problems of Adjustment—Society, Job, and Family."

The numerous contemporary illustrations and practical applications of psychological principles are likely to be interesting and useful for high school students. Despite the practical orientation, the material presented is drawn largely from experimental sources. The interpretation of this material is objective but eclectic. References to sources are omitted since "for the students for whom this book is intended such source material would have slight practical value, if any."

As a text for use on the college level, even if the general course is conceived in terms of its contribution to general education, this book seems to be somewhat superficial. As a high school text, however, it should prove to be highly satisfactory since it is written on an elementary and practical level and, at the same time, maintains an admirable degree of objectivity and scientific integrity.

Reed College. FREDERICK A. COURTS.

GUTHRIE, EDWIN R., AND POWERS, FRANCIS F. *Educational psychology*. New York: Ronald Press, 1950. Pp. vi+530. $4.00.

This volume is designed as an introductory text in educational psychology addressed to prospective teachers. It stresses the psychology of learning and its applications to classroom problems.

Its special quality arises from the fact that is is dominated by Guthrie's theory of learning by association. Stimuli are goads to trial-and-error activity. Eventually the cat or the child finds the response which permits escape from an unpleasant situation. This correct response gets learned because it is the last of the series of responses and thus is protected from unlearning.

> When a child enters school he confronts an enormous number of new skills which must be learned by trial and error. . . . He must learn how to sit rela-

tively still, how to remain relatively quiet. . . . In a large number of these skills, the motivating agent is the teacher. She is the goad that keeps prodding the child into activity until he achieves the required result (p. 120).

Guthrie's theory of learning has the respectability of being one of several recognized theories. The exposition of other theories is rather good but decidedly condescending. The implications of this theory for classroom teaching, as indicated by the above quotation, seem to me to be unfortunate.

Only rarely does another point of view emerge. Chapter XII begins with these lines: "The child is a whole and entire organism. He exists in a world of situations which are complex stimulus patterns. To these situations he responds as an entire organism." Chapter XIII begins, "A modern program of education calls for rich and varied learning experiences for every child." The promise of these words is not fulfilled.

There is much solid material in this text. For the most part the exposition is very clear and well organized. There are many homely references to actual classroom problems. There is a wealth of detail and of practical, common-sense advice.

The authors assert (p. 112) that different learning theories amount to much the same thing from the point of view of practical educational psychology. Their text is a demonstration to the contrary. The prospective teacher who absorbs this volume will be a different and, I fear, a poorer teacher of children.

FRANK K. SHUTTLEWORTH.

The City College of New York.

ANASTASI, ANNE, AND FOLEY, JOHN P., JR. *Differential psychology.* (Rev. Ed.) New York: Macmillan, 1949. Pp. xv+894. $5.50.

This revised edition has been greatly enlarged and improved over the original text. Extensive material from biology, child development, and anthropology has been added, as well as many new citations from the psychological literature, which has grown so extensively during the 12 years between the editions. While the general outline of the book remains the same, there is marked improvement in the organization of many chapters, in the use of graphs and tables, and in the general format.

In their preface, the authors propose to place major emphasis on methodological problems and basic concepts. This they have done. However, for a subject so involved with psychological theory, this text fails to present a clear-cut statement of principles and postulates which the authors obviously have assumed before undertaking their discussion of the literature. They attempt such a statement in the first four chapters but, in this reviewer's opinion, have failed to be consistent. After discussing and seeming to accept the "interaction viewpoint" concerning heredity and environment, the authors give little recognition to this interdependence. They carry forward habits of thinking which reflect behaviorism much more clearly than the "field theory" which is often implied by the interaction viewpoint. Throughout, the authors' sympathies are clearly with environmental determinism.

The authors' inconsistency in criticism appears most clearly in discussion of studies establishing differences between groups. The authors seem eager to discount studies which have established racial, national group, or sex differences unless the demonstrated superiority of one group in measured abilities can be associated with obvious environmental advantages, as in Klineberg's studies. With such studies, as with studies which report no differences, the text finds little fault. In other studies, the sampling is usually attacked, or if intrinsic or native differences are implied, alternative explanations in terms of learning are advanced. (One possible exception may occur in the fact that while there are lengthy discussions of how the feebleminded child's low IQ might result from environmental influences and hence be a social product, the child prodigy is accepted as a sort of psychological curiosity, admittedly valuable, without any searching for the social or environmental factors which produced him!)

Studies are criticized in detail which attempt to allow for environmental advantages of one group and yet do find group differences (e.g., Tanser, Bruce). When behavior changes occur in the *absence* of special environmental advantages, as specifically the changes which occur in loss of skill in a practiced group during an interval of no practice, then the authors feel it necessary to advance several hypotheses as to why this loss could occur. (Obviously it shouldn't!) Where studies frankly avow a cultural origin of the differences observed, such as culturally-determined differences in expressive movements, the studies are accepted with little criticism of methodology or with no mention of alternative explanations. One of the few exceptions to this is found on page 652 in a reference to girls' "demonstrated superiority in linguistic aptitude," which is accepted without particular debate.

The authors reject "growth" as having much to do with the intellectual and emotional development of the older child, though they admit that growth may play a major role in the behavioral development of the very young child (p. 265 ff.). They do not cite, however, the difficulty of an individual's later escaping from the impact of this early growth once it has set or sensitized the learner toward certain classes of stimuli. At one point the authors hold that a psychological growth curve is "at best a practice curve obtained in the absence of controlled conditions" (p. 278). The authors then assert that such curves can vary in different cultures. This reviewer gained the impression that the authors believe there is nothing particularly definitive or deterministic about the learning process, since learning curves are so subject to cultural vicissitudes. Yet an individual is the creature of his developmental *or* of his learning processes, and the cumulative significance of either process for the individual cannot be minimized.

The book is very complete in literature cited and in the coverage of its topic. While it is a little ponderous for undergraduates in psychology or education, it certainly is well suited to the advanced critical student. It deserves wide usage. Possibly the inconsistencies in criticism and interpretation which occur add to its merit for use with students who have some background in literature and in research method. It becomes a provocative source of discussion, where a more consistent book would be only doctrinaire.

DALE B. HARRIS.

Institute of Child Welfare, University of Minnesota.

HOPPOCK, ROBERT. *Group guidance: Principles, techniques, and evaluation.* New York: McGraw-Hill, 1950. Pp. xiv+393. $3.75.

This book is a practical handbook for guidance workers who plan to include group guidance procedures as a part of their total guidance program. The author avoids theoretical issues and makes the book consist primarily of clear, concise, practical suggestions regarding the content, techniques, and evaluation of group guidance. The content of the book is limited to the use of group guidance in orientation, educational guidance, and vocational guidance; other areas in which group guidance methods could be used are deliberately omitted. He makes it clear that the book is intended for teachers and school or college guidance workers but not intended for clinical psychologists and personal adjustment counselors. However, many of the techniques of group guidance and of evaluation of guidance could be utilized in group psychotherapy.

The book has four major parts: principles, techniques, evaluation, and appendixes. Part I, "Principles," discusses such practical questions as: Who should get it? Should it be required? Who should supervise it? What should it include? These questions are discussed only from the point of view of the use of group guidance within a school program.

Part II, "Techniques," outlines the purpose, preparation, and procedures of more than twenty techniques of group guidance, including group visits, student surveys, case conferences, visual aids, panel discussions, dramatizations, and others. The various techniques are concisely outlined and illustrated by summaries of articles from professional journals and other sources.

Part III, "Evaluation," stresses the need for objective evaluation of the outcomes of guidance. It consists primarily of brief summaries of twenty research studies, thus presenting specific examples of evaluative techniques, the variety of criteria utilized in evaluation, and the many problems involved in attempting to evaluate guidance.

Part IV, "Appendixes," is the longest of the four parts. It contains primarily verbatim records of demonstrations of many of the group guidance techniques which are outlined in Part II and helps to clarify the points stated in the text.

The author is to be commended for his frankness, sincerity, and modesty. For example, in contrast with some other writers of guidance books, he states clearly that he does not consider group guidance as *the* guidance technique but regards it as effective only when coordinated with other guidance techniques. Also, he states that "it would be odd if future research did not reveal some of his methods and conclusions to be erroneous." However, persons who plan to use or who already are using group techniques will find the book a useful and practical guide for organizing their programs, selecting and applying their techniques, and for evaluating their work objectively.

University of Redlands. WILBUR S. GREGORY.

PLANT, JAMES S. *The envelope: A study of the impact of the world upon the child.* New York: Commonwealth Fund, 1950. Pp. xi+299. $3.00.

It is a bit odd to read this last work by the author of *Personality and the Cultural Pattern.* The earlier work was a major attempt at synthesis and was the first clear expression of a movement which today incorporates psychologists, sociologists, anthropologists, and psychiatrists. The movement is known by the convenient, if ugly, expression, "Culture and Personality." The present work has none of the scope and depth of the theoretical work which it follows by some 13 years. Indeed it seems miscellaneous. But it only *seems* so. It is written around 21 problems, each of which takes a separate chapter. Illustrations of these problems are: security, authority, feeling of difference, intelligence, reaction to failure, minority race problems, narcissism, and the meaning of meaning. Each topic is discussed as an entity, but to it is applied very effectively the earlier-developed notions. The outcome is stimulating. The rich experience and insight of the clinician has been joined to the tolerance and humility of a person who sees and understands the complexity of society. There is no glibness, no false hopes; but there is no annoyance, either, at the reality picture of a dependent organism trying to adjust to shifting, inordinate demands and requirements. For him man is a temporary solution confronted by problems and challenges to these earlier adjustments.

The major theme, the "Envelope," refers to a large, integrated area of personality which interprets the social and physical world to the organism and *through* which he communicates himself to others. The Envelope is comparable to such terms as autism and motivated perception, though it is not, of course, identical with them. All such terms stress that there are various selective processes at work which change, distort, and vary the "input" to make it congruent with the person's needs and, therefore, to protect him. Where Plant's special sensitivity (and sensibility) comes in is in his handling of the interrelationships between various needs. He sees them as changing, as sometimes not what they seem to be, as arising, functioning, and disappearing in a matrix of social and "accidental" variables. And by "needs" he does not mean basic physiological processes alone; he refers to the requirements which one finds in the clinical examination of problem children. It is because of this that he can think of relating to personality problems which arise in school situations the "problems arising out of the fact that steam cannot be efficiently transported . . . that people have to move to this source of power. Electricity and chemical sources, on the other hand, are decentralizing . . . and their development will mean other shifts in social pressures on the individual." In his hands such observations make clearer the job of the clinician and the difficulties he will face.

The only weakness in the volume is the lack of concern with learning as a process. Given motivation and a problem, the behavior equipment seems always to be available and ready. But within the context of clinical problems perhaps the weakness is not Plant's but of those to whom he would look for aid in their solution. In any event it is the kind of book which one reads, enjoys, and learns from.

RICHARD A. LITTMAN.

University of Oregon.

EDWARDS, ALLEN L. *Experimental design in psychological research.* New York: Rinehart, 1950. Pp. xiv+446. $5.00.

This is an excellent book which will be of great value as a text and reference. It does not, however, have the content which some readers would expect from the title. Its content is better indicated by the following statement from the preface:

> This book attempts to present to the student in psychology, education, sociology, and the behavioral sciences, some of the newer developments in statistical analysis, particularly with respect to small-sample theory, as they relate to problems of research and experimentation in these fields.

The objective expressed in this statement is attained in admirable fashion. The author presents a variety of methods of statistical analysis. They are primarily the ones that the experimenting psychologist today is apt to find most frequently useful, and include some that are not widely known to those who would find them useful. He organizes his presentation primarily around types of research design, discussing in each case the analyses appropriate to the design. The presentation is admirably clear. In my opinion, this book would provide for very able students a first-rate introductory textbook on statistics (though, since it is obviously not intended for introductory use, some supplementation would be needed). Where the student cannot be given the advantage of such an introduction, the book should be highly useful as a text for an advanced course in statistics. Where a combined course is given on statistics and experimental method, this book provides a practically ideal text if the students are up to it. For advanced students and psychologists the book provides probably the best available general guide to the statistical analysis and evaluation of experimental data.

This book, despite its excellence, does not fully meet the need for a textbook on experimental design in psychology. To state this is not to criticize the author, since he apparently did not have this as his *main* aim in writing the book. In view of the title which the book bears, however, it seems important to state why the book is not fully adequate in relation to this need. Some of the reasons follow:

1. From the point of view of this need, the great emphasis on statistical analysis is distracting. For the student whose principal aim is to learn about research design, matters of statistical analysis should be treated in proper perspective as a tool, rather than be the main subject of discussion, if he is not to be diverted from his aim.

2. From the same point of view, there are some important omissions. The most glaring one is the absence of explicit general discussion of the frequently used design in which a subject is used as his own control and there is hence need to counterbalance experimental treatments and possible temporal effects. In the chapter on "Experimental Designs Involving Repeated Measurements of the Same Subjects," this design is not even mentioned. It is discussed only in the chapter on the Latin square, and there primarily as illustrative of the use of the Latin square rather than as a design involving some distinct problems of its own.

3. On problems of design in contrast with analysis, the author does not write at much length or with as great understanding and awareness of problems.

4. It appears to be something of a defect that the author has not introduced much of the terminology which has been developed by Fisher and others for aspects of research design. Psychologists and their students will presumably need to continue to consult the writings of these men. It would facilitate this reference if terms such as "randomized block" and "split plot" design had already been encountered in a book of this sort. These aspects of design are treated here, and no other single labels are offered for them as more appropriate in psychology. To have provided any label would have facilitated the student's getting a useful concept; to have provided the Fisherian label would have the added advantage of increasing the student's scientific literacy.

Despite these objections, the book represents a definite gain to the psychological literature, even on the subject of experimental design. It doesn't do as much as one could hope, but it does something very valuable which has not been done before.

I have a few comments on specific parts of the book, which may be of use to some of its readers. They follow:

1. The author does a service in bringing into the psychological literature a useful convention of referring to one-tailed tests by the term *significance point* and two-tailed tests by the term *significance level*. The clarity and consistency with which this distinction is used here should help in eliminating a source of unnecessary confusion. It should be noted, however, that the distinction as used here may lead the student into fresh confusion at one point—the interpretation of F when used to test the significance of the difference between two means. The usual use of the F tables for this problem yields a significance *point* with respect to the distribution of F, but this is equivalent to a significance *level* with respect to the difference between the means (since a high value of F may be produced by a difference in either direction between the two means). This fact is not taken account of by the author, and the student who adopts Edwards' terminology will be puzzled when he finds that we do not then, as Edwards claims, "arrive at exactly the same conclusion that we reached by means of the t test" (p. 184). We would, in fact, arrive at conclusions which differed by a factor of 2. The confusion could be cleared up by applying the term *significance level* to the one-tailed F test in this particular case; if one wants for this case a *significance point* which will correspond to that obtained through t, it will be necessary to divide the tabled p value by 2.

2. The author's italicized statement on pages 216–217 is too sweeping and might well lead a student to make an incorrect test of significance in analyzing some of the designs discussed in chapters 15 and 16.

3. There seems to be a slightly misplaced emphasis in the author's suggestion (as on p. 5) that it is important to try to get *continuous* scales of measurement. Surely for almost all purposes we would gladly settle for *discrete* scales; there are other scale properties much more important to worry about than this one.

4. The author's statement of the meaning of triple interactions (pp. 227–228) is not at all clear. A person who does not already know their meaning is likely to understand from it that they simply confirm the existence of several first-order interactions.

5. The author's statement on page 257 that "We would thus be led to the false conclusion, in this analysis, that none of the main variables is significant," is misleading. If the hypothetical experiment referred to had been done in the way described here, this would be the true and appropriate conclusion.

6. It is unfortunate that Edwards' discussion of systematic Latin squares (pp. 307–310) is so inconclusive. He gives the impression that methodological research is the only source of usable knowledge about the bias of systematic squares, whereas purely formal knowledge might well be of value.

7. The value of the book is increased by the great diversity of examples described, both in the main text and in the problem section at the end of each chapter, by the numbered list of formulas printed just before the Appendix, and by the well-selected tables contained in the Appendix.

IRVIN L. CHILD

Yale University.

Cochran, William G., & Cox, Gertrude M. *Experimental designs.* New York: Wiley, 1950. Pp. ix+454. $5.75.

The authors attempt to ". . . describe in some detail the most useful of the designs which have been developed, with accompanying plans and an account of the experimental situations for which each design is most suitable" (p. v). They succeed admirably in achieving this and other objectives. It is assumed that the reader has had more than a casual exposure to analysis of variance, upon which virtually all of the analyses depend. There is no emphasis on non-parametric situations and specialized tests.

The first three chapters deal with: (1) the contribution of statistics to experimentation, and initial steps in planning experiments; (2) methods of increasing the accuracy of experiments, including replication, refinement of measures, design, and grouping data; and (3) notes on analysis of variance. The first two chapters contain much wisdom, stated with such simple humility that the presentation attains a homely beauty. In chapter III easy approaches are made to difficult topics. Likelihood ratio tests are gently introduced. Mundane but difficult matters such as accuracy of computations are treated with care born of long experience with real data.

Chapter IV deals with randomized designs, single, double, and triple groupings, and chapter V covers factorial experiments. Each design is discussed under: description, randomization, statistical analysis, standard errors for *t*'s, missing data, and efficiency. Lists of references and detailed experimental plans appear at the end of each chapter. The frequent numerical examples in the text encourage penetrating analyses, far beyond the initial *F*'s of the analyses of variance. Little special comment is made on the necessarily a posteriori nature of some of these analyses. Confounding is systematically dealt with in chapter VI, and summary tables of available confounded designs are conveniently given along with indices of the locus and degree of reduction in efficiency due to confounding. Chapters VII–XIII deal with confounding and balanced designs: split-plots, quasi-latin squares, balanced and partially balanced incomplete blocks, lattices, lattice squares, and incomplete latin squares. These chapters (pp. 218–390) contain numerous examples, details of numerous plans or layouts, and indexes to these plans. The material is presented with little repetition and with abundant cross-reference, so that the careful reader can almost use the text in cookbook fashion.

In chapter XIV, the analysis of a series of experiments is attempted. The steps here advocated require experience and "nerve." Procedures to follow when the standard assumptions fail are outlined briefly, but the rationale is consistently omitted. The last chapter gives tables of random permutations of 9 and 16 numbers with instructions on their use, and the book concludes with a selected bibliography and the usual index.

Psychologists should profit immensely from this fine book. Long before they heard of Fisher, psychological researchers used orthogonality and balancing procedures in their experiments. Recently they have learned something about the proper statistical analysis of their designs. Another less intuitively obvious set of procedures is available in confounding and balancing, and Cochran and Cox make these more available to the psychologist than do any previous authors. The reviewer predicts that extensive use will be made of the balanced designs—chiefly of the squares, latin, lattice, Youden, etc. These designs will be especially useful when a multiplicity of treatments cannot all be given to the same experimental subject because of time, fatigue, transfer effects, etc. *Experimental Designs* opens the door to what has been a relatively inaccessible area to most of us, and the authors merit our thanks. It is the reviewer's impression that the more the reader understands the basic principles of design, the more he will applaud the achievement represented by this book.

David A. Grant.

University of Wisconsin.

Harris, J. D. *Some relations between vision and audition.* Springfield, Ill.: Charles C Thomas, 1950. Pp. v+56. $1.50.

When sense modalities are being compared, relations of difference are more easily established than those of similarity. But many sensory phenomena transcend modal boundaries, and some, such as adaptation, appear in all sense departments. This little book, consisting of a dozen or so brief essays (the majority less than a thousand words in length), stresses the functional resemblances between vision and audition as a start "toward arriving at general principles of organization and theory of the whole sensorium."

The points of comparison of the two senses, not entirely systematically selected, themselves exemplify the difficulties inherent in making intermodal comparisons. The topics are: absolute and differential sensitivity, the relative ranges of intensities mediated, wave frequency as a sensation determinant, relative efficiencies as energy integrators, ways in which sensations develop and decay, responses to regularly interrupted stimuli (flicker vs. flutter), the question of bilateral interaction, events in single nerve fibers, central and peripheral determinants of acuity, quantum considerations, and intersensory phenomena.

The experimental evidences cited are quite commodiously documented (a bibliography of over 100 titles). Still, there are a few topics in connection with which one could wish that the visual literature had been appealed to in as competent a manner as was the auditory.

Frank A. Geldard.

University of Virginia.

Pastore, Nicholas. *The nature-nurture controversy.* New York: King's Crown Press, Columbia Univ., 1949. Pp. xvi+213. $3.25.

Pastore strongly believes that conservatives are hereditarians and that liberals are environmentalists. To support this thesis he examines the writings of distinguished hereditarians and environmentalists and finds eleven out of twelve hereditarians to be political conservatives, and eleven out of twelve environmentalists to be political liberals. Pastore feels that the political attitudes determine the scientific views. To quote (p. 181): "Different scientists are differently affected by their political allegiances. . . . The nature-nurture controversy, qua controversy, has been sustained by the conflicting social philosophies of the scientists. . . . It is probable, however, that in most cases the scientists were not aware of the specific impact of their political loyalties upon their scientific thinking."

Eight specific comments can be made:

1. The selection of cases was made by Pastore, who admits that the sample is not statistically valid.

2. The other regression equation might well have been used by selecting liberals and conservatives and seeking their views on heredity and environment.

3. Since no additional judges were used, the results present Pastore's image of the views studied, not a consensus. He says (p. 17) that the cited quotations are intended to typify a particular point of view, but adds later that they are not meant to be representative of the total expression of the person quoted.

4. The effect of change with chronological age is neglected. Is the Watson of *Comparative Psychology* the same as the Watson of *Psychological Care of the Infant and Child*?

5. The era or social climate in which the person wrote is not controlled. In their own time, some whom Pastore labels as conservatives may have been liberal. All those studied lived after 1822, in the period of the greatest social change and the greatest advance in science. The hereditarians were all born prior to 1888, whereas four of the twelve environmentalists were born later than 1890. In Pastore's own classification of the periods in which the persons studied were influential (pp. 17–18), ten of the hereditarians start with 1900 as compared with four of the environmentalists, while two of the hereditarians start with 1919 as compared with eight for the environmentalists. What Pastore may be studying is the progressive modification of thinking as the complexity of the problem of heredity and environment has become more apparent with time.

6. In order to equate the data upon which the opinions studied are based, more control should be exercised over the intellectual and scholarly background of the persons studied. According to Pastore's classification, the hereditarians consist of six psychologists, three geneticists, two biologists, and one statistician, whereas the environmentalists consist of four psychologists, one anthropologist, three geneticists, one biologist, two sociologists, and one educator. This suggests that the type of data with which the person has had experience may determine

attitudes. Studies of nature-nurture may possibly determine political and social attitudes.

7. The objective effects of the work of the particular persons upon society is not discussed. Some who are listed as hereditarians have had a much greater impact in creating better atmospheres and good lives for human beings than some who are listed as environmentalists.

8. No attention is given to intermediate positions in either variable studied; a scientist must be classed either as an hereditarian or an environmentalist, as a liberal or a conservative. Many of the statements quoted, however, indicate intermediate positions. This is particularly true of the geneticists, whether or not they are placed in the conservative or liberal group.

The semantic problem with which Pastore struggles is a difficult one. But his solution is too easy. If his essential thesis is correct, namely, that a person's scientific views are a product of his political attitudes, in justice to his readers Pastore should state his own political attitudes, since they may be determining his own analysis of the data. From the book as a whole, one gets the feeling that Pastore believes that hereditarianism, conservatism, and badness are identical, while environmentalism, liberalism, and goodness are identical. This reminds one of the Stendler finding that first-grade children confuse beautiful, rich, and good and that as they mature, the concepts are differentiated.

JOHN E. ANDERSON.

Institute of Child Welfare, University of Minnesota.

HURLOCK, ELIZABETH B. *Child development.* (2nd Ed.) New York: McGraw-Hill, 1950. Pp. xvi+669. $4.50.

This book represents an extensive revision of a college text first published in 1942. The most significant change has been the abandonment of the longitudinal approach of the earlier work, in keeping with what the author states is "the recent trend in many colleges and universities to separate the study of child development from that of adolescence and to present the two in different courses" (p. ix). Because of the increasingly large amounts of research in both fields, she continues, it is impossible to cover the whole span of development in a single book. Therefore, all material pertaining to adolescence has been removed from this edition, and three new chapters have been added on childhood interests, family relationships, and personality.

The first two chapters present, respectively, a brief sketch of the history of child psycholgy and of the principles of development. Including the new material, the remaining thirteen chapters deal with the usual topics of child psychology, such as prenatal development, the newborn, physical growth, motor development, speech, play, the development of understanding, and emotional, social, and moral development. A bibliography is appended, comprised of some 1,250 items distributed over the last thirty years. In the relatively few instances where explicit references are not given, statements are made which presumably reflect the general literature or the author's own views.

While the order of the chapters is fairly typical of texts of this sort, the basis for the arrangement of material within a chapter is not as easy to grasp. Furthermore, the inclusion of some systematic generalizations with reference to the major aspects of human behavior discussed and to their *development* would have added to the generally excellent presentation of the enumerated studies and would possibly have made it easier for the student to synthesize the citation of so many specific pieces of research. However, the liberal use of pictures, tables, graphs, charts, and other illustrative materials makes this text not only interesting but also a very readable elementary handbook in its area.

In terms of age range and subject matter this book is even more, as Dennis observed (*Psychol. Bull.*, 1943, **40**, 536–537) of the earlier work, "a college textbook in child psychology." It is one, however, that is clearly written, and it does present the student with well-digested, documented, and illustrated summaries of research from a vast literature on a wide variety of topics in the field of child psychology.

Fisk University. S. O. ROBERTS.

VON BONIN, GERHARDT. *Essay on the cerebral cortex.* Springfield, Ill.: Charles C Thomas, 1950. Pp. xiv+150. $3.75.

Written by an active and leading student of neuroanatomy, this essay on the cerebral cortex emphasizes structural differentiations and relations as determined by anatomical and histological techniques. Physiological methods, behavioral data, and psychological analysis are accorded only nominal mention. The essay is divided into seven chapters: "Historical Introduction" (12 pp.), "Theoretical and Phylogenetic Introduction" (15 pp.), "General Morphology of the Cortex" (33 pp.), "Sensation" (33 pp.), "Action" (18 pp.), "Prediction" (7 pp.), "Emotion" (8 pp.). There are a few original suggestions—notably in regard to the extrapyramidal system and the somesthetic pathways—but in general the author maintains a conservative position and a modest, non-dogmatic attitude.

From the author's basic view that the output of the brain "is a definable function of its input," it follows that sensation and action "are obvious chapter headings." But this concept, he points out, is inadequate in reference to the major portion of cortical activity, since man developed languages "based on a 'symbolific' function . . . which cannot be considered an output in the sense in which a simple muscular movement . . . is called an 'output.'" The central difficulty, of course, lies in the fact that psychological analysis, which "has to proceed independently of anatomy and physiology," has made little progress in defining the factors or functions which neurology is to localize. Furthermore, "we have no right to expect that all the factors will be localizable in the sense in which classical tradition used that term." However, since "both the koniocortex of the primary sensory areas and the agranular cortex of the precentral motor cortex must remain unencumbered by things past in order to function efficiently," memory and prediction are localized, by elimination, in the homotypical cortex. Looking forward as well as looking backward "must be based on the possibility of interchanging spatial and temporal relations."

In respect to general principles of neurological investigation, von Bonin affirms that "the function of the cerebral cortex cannot be explained by one principle. . . . Two systems are of paramount importance, the 'random' intracortical and the 'ordered' corticocortical system. . . . Within the (macroscopic) corticocortical system the causal laws of macroscopic physical events, as we know them from mechanics, will hold good; within the (microscopic) intracortical system the laws will be statistical as in thermodynamics or like those which the physicists develop for 'fields,' as in electromagnetism."

The virtues of this little volume are a forceful demonstration of the driving force behind the "conviction that structure should be understood as an enduring order impressed upon a flow of energy." Its deficiencies perhaps reflect more the magnitude and inherent difficulties of the problem than any specific inadequacies of the present author.

HENRY W. NISSEN.

Yale University.

PARTEN, MILDRED. *Surveys, polls and samples: Practical procedures.* New York: Harper, 1950. Pp. 624. $5.00.

This book represents a comprehensive treatment of a field which, in spite of a considerable body of work, is still more art than science. On the whole, the author succeeds admirably in giving a balanced picture of the survey field, with emphasis on "practical procedures." The chapters on planning, methods of securing information, construction of the schedule or questionnaire, interview and mail questionnaire procedures, sources of bias and processing (editing, coding, and tabulation) provide an excellent (and practical) guide to these phases of survey operations. They should prove useful both as a text for courses in survey procedures and as a "check list" for survey directors looking for weak spots in their own operations and suggestions for modifications. The value of the book is considerably enhanced by an extensive bibliography.

On the debit side are the unsatisfactory treatment of sampling and the unqualified presentation as "fact" of many items which are, at best, "opinion." An example of the latter defect is the statement on page 161, "processing the data from booklets is awkward and time-consuming. For efficient processing it is almost essential that the information be transcribed to a form better adapted to office procedures." No evidence for this conclusion is presented, and the reviewer knows of no conclusive demonstration that processing a well-designed booklet questionnaire is less efficient than processing any other type of questionnaire which contains the same amount of information.

Many of the assertions in the book are undoubtedly true. Most of them have not as yet been experimentally verified. Many readers will fail to grasp what is, I am sure, clear to the author of this book: there is considerable experience and extensive "folklore" in the survey field but only a meager amount of scientifically verified fact. The book has done a painstaking and admirable job of collecting a wealth of experience from the individuals and organizations in the survey field. The reader must continually bear in mind that he is reading a summary of accumulated experience and that uncritical application of the experience of others may be dangerous.

The treatment of sampling is quite disappointing. If the book is used as a textbook in a course on surveys, the instructor will do well to omit these chapters and draw on other sources. There is some excellent material presented (particularly in the chapter "Procedures for Drawing Samples"). On page 64, there is set forth an excellent statement of the aims of good sampling and the reviewer heartily endorses the statement (p. 403) that "research in the social sciences is still at a stage where biased errors are likely to be twenty to a hundred times as great as random errors unless the survey and sampling are very carefully designed and executed." In spite of these good features the sampling discussion is both confused and inaccurate.

Probably the most serious defect of the sampling discussion is the presentation of material based on the theory of unrestricted random sampling with the implication that it applies (at least reasonably well) to almost any type of sampling. While there are some cautions on this point, the general effect is to reinforce the careless procedures fostered by the inadequacies of most of the statistical textbooks used by psychologists. The discussion of sampling also perpetuates the fallacious identification of quota sampling with stratified sampling.

One point which will be forcibly driven home by a critical reading of this book is the inadequacy of published data about survey costs. The author, quite correctly, stresses the importance of cost considerations in planning surveys. While the discussion of survey costs does not explicitly indicate the paucity of cost data, the information presented speaks for itself. In passing, the reviewer questions the accuracy of the statement that "The Bureau of the Census has found payment per schedule the most satisfactory because enumerators have to make their calls when there is a chance of finding people at home."

There are other minor inaccuracies (and a few major ones) in this volume. These defects may be unavoidable in a book of this scope. In spite of the defects, there is a wealth of valuable advice for the research worker who reads the book carefully *and* critically.

Eli S. Marks.

U. S. Bureau of the Census.

Freeman, Frank S. *Theory and practice of psychological testing.* New York: Holt, 1950. Pp. xxi+518. $3.50.

The purpose of this book is stated clearly in the preface: "In the developing scientific field of psychological testing, there is recurrent need for periodic presentation of comprehensive descriptions of tests. This volume is intended to meet such a need." The major portion of the book is, accordingly, devoted to descriptions of a large number of tests. Considerable space (54 pp.) is devoted to the Binet scales and their many revisions. Other chapters are devoted to "Semi-Performance and Performance Scales," "Nonverbal Group Scales of Mental Ability," "Verbal and Mixed Group Scales of Mental Ability," "Scales for Infants and Preschool Children," "Aptitude Tests," "Tests of Eductional Achievement," "Personality Inventories and Rating Scales," "Projective Methods," and "Situational Tests." In addition to the enumerations and descriptions of tests provided in the chapters listed above, the book includes chapters on "Basic Principles: Theoretical and Clinical," "Statistics of Psychological Testing," "Definitions and Analyses of Intelligence," "Intelligence Tests as Clinical Instruments," and "Applications and Problems." The last is a sort of catch-all for brief presentation of topics like the nature-nurture problem and racial and national differences.

The book is well written and should be a very useful textbook for courses in educational and psychological testing. The material included seemed to the reviewer to be satisfactorily accurate, although the tone or emphasis sometimes struck him as inappropriate. However, this is to be expected as a result of differences in points of view and is not a criticism of the book. The author is considerably more lenient with respect to the tests he discusses than the reviewer would be. It is entirely possible that in an elementary text materials that deserve severe criticism should simply be ignored; otherwise, students may become confused. The reviewer would like to have seen somewhat greater emphasis given to the flaws and limitations of tests such as the Bellevue Scales and the Chicago Tests of Primary Mental Abilities. As it is, the misinformation and speculation in Wechsler's *Measurement of Adult Intelligence* is never called to the student's attention. And the spuriously high reliability coefficients reported in the manual for the Chicago Tests of Primary Mental Abilities are accepted at face value. Unfortunately, a competing set of tests, the Differential Aptitude Tests (published by the Psychological Corporation), for which parallel-forms reliability coefficients are provided, is not mentioned in the book.

It is gratifying to note that the author is cautious in accepting the claims made for tests based on the results of factor analyses. Although the reviewer is firmly convinced of the value of factorial techniques, he is worried by the uses to which these are sometimes put.

A couple of what seem to the reviewer to be non sequiturs were noticed. Both are of minor importance.

1. In the footnote on page 314, we find the statement: "But so far as AQ's above 100 are concerned, they are illogical and psychologically unsound, for actually no one can achieve above his capacity." It is no doubt desirable to point out the dangers of the AQ, but it seems a bit strong to indicate that all AQ's above 100 are illogical and psychologically unsound. Though it may be impossible, by definition, for anyone to achieve above his ultimate capacity in a given field, this does not preclude our obtaining meaningful AQ's above 100. For example, a child may show as much capacity on a certain mental test as the average child does at age 10–0 but score as high on a test of general science as the average child does at age 12–0 because of intense interest in the field and favorable opportunities to follow his interest. An AQ above 100 for this child would be quite meaningful.

2. On page 173, the author mentions a correlation coefficient between scores derived from the Pintner-Paterson Scale and the Stanford-Binet Scale in a group of gifted children, and concludes: "Furthermore, the very low coefficient of .23 suggests, also, that the performance scale is particularly inadequate for differentiating between performance levels of gifted children." To the reviewer the coefficient suggests that the distribution of scores for each test was probably greatly restricted. In any event, it would be possible to have two tests yielding uncorrelated scores that would each make excellent discrimination among gifted children.

Hunter College. FREDERICK B. DAVIS.

HEPNER, HARRY W. *Psychology applied to life and work.* (2nd Ed.) New York: Prentice-Hall, 1950. Pp. viii+724. $4.50.

In this thorough revision of Hepner's book, more material has been added, the structure has been tightened, and new chapters are included on such topics as interviewing, counseling, supervision, family problems, and group dynamics.

The section on industrial psychology represents perhaps the most outstanding part of the book. Although the individual worker's role in the industrial environment is thoroughly covered, emphasis is placed on the implications of the important group research findings, such as the Hawthorne investigations, the "group atmosphere" studies of Lewin, Lippitt, and White, the new reports from the University of Michigan's Survey Research Center, Leighton's work on morale, and the advances in sociometry and attitude measurement.

Most of Professor Hepner's many illustrations and examples are properly documented, but there are instances when a more critical evaluation of the illustrations would have been desirable. General Motors' "Why I like my job contest," for instance, is described without a critique of some of its sampling procedures and assumptions (pp. 471–473). A statement that "union members are more inclined to favor 'holding back' (of production) than are non-union workers" (p. 401) needs support in view of the contradictory evidence recently provided by Selekman.

Similarly, the section on the adjustment concept contains some conclusions based on clinical insight which could have been more profitably supported by experimental evidence. The statement "to understand an adult we must understand *how he felt* as a child" (p. 113) requires examination in the light of findings of the field theoreticians. The statement that "the strength of a nation . . . depends mainly upon the strength of personality of its people" (p. 45) introduces a vaguely mystical conceptualization. The belief that "the human race has progressed because heroic individuals have used positive adjustment in the face of adverse situations" (p. 44) is not particularly strengthened by the case history of a life-term prisoner who was able to make good in the advertising business.

The above should not detract from the over-all excellence of the book. In his revision, Professor Hepner has improved an already good book, which no doubt will again enjoy wide popularity among teachers and students alike.

IRVING R. WESCHLER.

University of California, Los Angeles.

FROMM-REICHMANN, FRIEDA. *Principles of intensive psychotherapy.* Chicago: University of Chicago Press, 1950. Pp. xv+246. $3.75.

This book has many distinctive and admirable qualities. It is written with an almost epigrammatic conciseness and clarity. It is calm and objective in mood yet definite, confident, and provocative. Indeed, it is a daring book in that it describes many technical innovations and serenely disregards entrenched dogma and doctrine.

Throughout the book one senses the long and varied experience, the wisdom, and the expertness of the author. Useful observations are made regarding almost every type of personality deviation which a psychotherapist is likely to encounter; but the author is particularly interested in the intensive treatment of psychotics, and it is here that some of her most trenchant and original discussions are to be found. Outstanding in this connection is the author's demonstrated premise that psychotics are treatable on the basis of the same underlying principles as are applicable in the neuroses and that the "security operations," or symptoms, of these two classes of patients are not basically different. To many readers the discussion of the qualifications and vicissitudes of the therapist as a person will also be of special interest and value.

The author reflects, and has importantly contributed to, the thinking of the late Dr. Harry Stack Sullivan and other psychiatrists, psychologists, and social scientists who have been connected with Chestnut Lodge Sanitarium, the Washington School of Psychiatry, and the William Alanson White Institute of Psychiatry in New York. In keeping with the general position of the "Washington School," Dr. Fromm-Reichmann accepts many of the basic tenets of Sigmund Freud: the reality of repression (Sullivan preferred the term "dissociation"), the anxiety theory of symptom formation, the therapeutic significance and utility of transference ("parataxic distortion" of interpersonal relations), and the premise that the primal pathogenic conflict is between cultural pressures for conformity to moral standards and the biologically given impulses of sex and aggression. On the other hand, Dr. Fromm-Reichmann, like many of her colleagues, swiftly dismisses the speculations of Freud regarding the "death instinct" and libido theory, expresses serious reservations regarding the classical Oedipus complex, and breaks with established traditions regarding many of the more routine features of psychoanalytic procedure.

This book is so simply written that beginners in psychotherapy will find it a useful and inspiring guide, and it is so replete with shrewd insights and general clinical sophistication that few experts will fail to enjoy and profit from it. From the standpoint of the reviewer, the book has only two major defects. In what we may suppose was the author's wish to give a semantically streamlined and highly operational account of psychotherapy, the book at times descends to a level which is slightly reminiscent of cookbook rules. This shortcoming will be readily apparent. The other one is more difficult to identify. It will probably be first perceived as a lack of theory and conceptual integration. Many of the author's observations are laid end to end without being tied together in any systematic way. But as one thinks further about the book, it will become evident that this weakness is related, somewhat adventitiously, to the book's greatest strength. As already indicated, the author follows the lead of Sullivan in discarding most of the biological baggage that psychiatry and even psychoanalysis (in a little different guise) has traditionally carried. Dr. Fromm-Reichmann places the central problems of psychotherapy squarely in the context of interpersonal relations, where they assuredly belong. However, the author's case would have been materially strengthened if she had not also disregarded certain of the constructs, or intervening variables, which form so important a part of Freud's later theorizing. From a formally theoretical standpoint, the book as it stands represents a somewhat naive form of social behaviorism; but this is a defect which could be easily remedied and is more than compensated by the book's many virtues and strengths.

University of Illinois. O. H. Mowrer.

LEFEVER, D. WELTY, TURRELL, ARCHIE M., & WEITZEL, HENRY I. *Principles and techniques of guidance.* (Rev. Ed.) New York: Ronald, 1950. Pp. vii+577. $4.25.

This ambitious volume of fifteen chapters is offered as a thoroughgoing revision of the first edition with the addition of two new approaches. First, while subscribing to the role of the guidance specialist, more attention is given the part played by the classroom teacher in the guidance program. Second, vocational guidance receives increased emphasis in the revised volume. The fifteen chapters are arranged into four parts. Part I deals with the basic concepts and principles of guidance; Part II with the group approach to guidance; Part III with individual guidance and counseling; and Part IV with evaluation and follow-up.

Contents of the book vary considerably among the chapters both in quality and detail of treatment. There is appropriate emphasis on the need for personalization and individualization of the curriculum. The usual techniques and tools of guidance are described and illustrated, some more thoroughly than others. The discussion of basic principles seems to the reviewer to be fundamentally sound, and there is an especially good chapter on tests. One chapter on selected guidance systems is timely, and will be helpful to those struggling with problems of organization.

The need for the guidance program, according to the authors, stems from the absence of an ideal system of education, and hence is recognized as a temporary expedient. Group approaches to guidance presently are therefore necessary. One chief function of the counselor is that of directing and supplementing the work of the teacher. It is not clear to the reviewer whether the authors would subscribe to the need for the counseling specialist under an ideal system of education.

Brief reference is made to nondirective counseling. It is recommended that "autocratic, prescriptive, and directive approaches" should be replaced by "some of the techniques" of the nondirective method. Just what of the techniques could be used without modification in the highly authoritative guidance setting presented by the authors, is not indicated. Elsewhere in the book it is suggested that techniques and methods used should lie somewhere between more highly structured and nondirective methods, which is hardly more clarifying.

While the authors regard the part-time or full-time counseling specialist as playing the double role of leader of the guidance program and specialist in dealing directly with certain adjustment problems, there is an absence of suggested framework within which the counseling specialist could coordinate the guidance efforts of classroom teachers and other staff members in meeting without duplication of effort the needs of the individual student. This, in the reviewer's opinion, is the greatest weakness of the book.

The guidance needs of exceptional students are recognized, and some of them are dealt with in detail. This discussion, however, has limited application to those of superior intelligence and special talents, to whom a total of only two and a half pages are devoted. Priority is given to students with various kinds of problems—recognition of a problem being necessary to set the guidance machinery in motion. The authors are not alone in failing to give sufficiently detailed attention to the needs of those who present no obvious problems, but whose superior capacities need direction into stimulating outlets. In spite of its limitations the book has much to offer the discriminating reader.

New York University. JOSEPH V. HANNA.

KENT, GRACE H. *Mental tests in clinics for children*. New York: D. Van Nostrand, 1950. Pp. ix+180. $2.45.

The author states that this book is intended for advanced students in clinical psychology. She states further that it was her wish "to present a point of view not generally recognized." This reviewer thinks that the author has achieved both goals with unusual success. This book will be of considerable help and interest to advanced students in clinical psychology, and the point of view, while not generally recognized, needs to be emphasized.

In chapter I, Dr. Kent introduces the problem of mental tests for children by considering the reasons for children's being referred to a clinic. The reasons discussed include mental retardation, reading difficulties, adoption suitability, and misconduct. The clinical examination is discussed in chapter II. The author emphasizes the need to make the examination appeal to the child, the use of praise, and the clinical observations.

Chapter III, "Sources of Error in Use of Tests," might be considered as the author's statement of rationale. The point of view is that undue emphasis and importance have been given to the quantitative aspects of mental testing. Diagnostic errors may derive from taking scores too literally. This emphasis is the major contribution of this book. For example, in chapter IV, "The Composite Scale Versus the Battery," the author states, "There is naturally a demand for the composite scale, on the ground that it can be presented more or less mechanically by a psychometrist of comparatively little training. It calls for more judgment on the examiner's part to select sub-tests suitable for a given subject than it requires to run through a rigidly prescribed series of tasks. It must not be forgotten, however, that the results of the mechanically presented examination are frequently misleading."

This reviewer finds the emphasis on the need to base a diagnosis on the complete appraisal of the child, rather than primarily on his scores, to be refreshing and stimulating. Some psychologists will not agree with this point of view, but Dr. Kent speaks authoritatively, and her discussion reveals a wealth of clinical experience from which her arguments derive. It is a practical and timely book. Clinical psychologists in service clinics and those responsible for training will find it helpful.

Northwestern University. HELMER R. MYKLEBUST.

HAHN, M. E., & MACLEAN, M. S. *General clinical counseling.* New York: McGraw-Hill, 1950. Pp. xi+375. $3.50.

Written as a text for graduate students preparing to become professional counselors, this book discusses the principles and functions of counseling and describes some of the basic counseling techniques. The authors' approach is problem-centered or, perhaps better, centered about students with problems, and the general role of counseling in education for personality development is given little attention. Accepting this limitation, *General Clinical Counseling* is a commendable text.

The first part of the book concerns itself with a definition of counselors and counseling. The problem orientation is obvious in this definition: "Clinical counseling, as the term is used in this book, is a process which takes place in a 1-to-1 relationship between an individual troubled by personal problems with which he has been unable to cope alone and a professional personnel worker whose training and experience qualify him to aid others to reach solutions to various types of personal difficulties." The 1-to-1 process is emphasized to the point that counseling involving two or more clients—for instance, marriage counseling—falls outside the definition of clinical counseling.

Included in this section is a discussion of the preparation desirable for counseling. Conspicuous through its absence is the statement issued by the National Vocational Guidance Association in 1949.

Next follow sections on principles of clinical counseling and the nature of various types of problems, including vocational problems and educational problems. Problems with broader implications are not included in this section nor discussed systematically in the book.

The major portion of the book is devoted to a discussion of the tools and techniques of counseling. These include counseling records, the interview, standardized tests, sociometric devices, occupational information, and statistics. It seemed to the reviewer that undue attention was given to the Minnesota Occupational Rating Scale. Although this scale has been found to be extremely useful in the training of counselors, its actual use in counseling interviews is relatively limited.

The final section of the book is devoted to a systematic case study. The case study would be of greater instructional value if it included actual excerpts from the interviews. Before interviews could be recorded and accurate typescripts obtained, a textual description of case work was the best that could be used for instructional purposes, but now that recording equipment is so easily available, samples of the actual interviews can be used most effectively.

The chief criticism that can be made of this text is that it looks backward rather than forward. Most of the discussion centers about what counseling has been, rather than what it could be. Much of the relevant research done in counseling, and certainly there has been relatively little, is omitted, and the authors apparently do not base their presentation on that research. Such pertinent research as that recently coming from the University of Chicago is not mentioned. The recent emphasis on the social context in which counseling takes place also fails to appear in this book, although the recent contributions of semantics and sociometry are recognized. The social significance of counseling is not explored.

Worth particular mention is the book's fine and lucid style of expression. It provides both interesting and pleasurable reading.

University of Minnesota. RALPH F. BERDIE.

LANDIS, CARNEY, & BOLLES, M. MARJORIE. *Textbook of abnormal psychology.* (2nd Ed.) New York: Macmillan, 1950. Pp. x+634. $5.00.

According to its authors, "The present revision was undertaken because of the vast amount of new research findings in the medical, social, and biological sciences during and since World War II." The writers have interwoven these recent findings into the fabric of classical abnormal psychology in such an excellent fashion that one cannot tell where the old leaves off and the new begins. As was true of the original edition, published just five years ago, this revision is certainly one of the best treatments of abnormal psychology available today. In some ways it is the best.

There appear to be only a few important studies and recent trends omitted in the revision, but in the reviewer's opinion certain such omissions are to be regretted. For example, in the discussion of typology the authors cite Kretschmer's classical types but do not mention the methodologically improved studies of Sheldon. The recent elaborate factorial studies of Cattell and Eysenck are not treated, even though factor analysis appears to be one of the more important methodological advancements in the field of abnormal psychology. The very recent cooperative endocrine studies of the Worcester group under the direction of Hoagland and Pincus are not mentioned, despite the fact that these studies offer one of the most promising attacks on all the psychoses. The significant aspects of role-taking as elaborated by Cameron are overlooked, as are the recent attempts to bring behavior disorders into the theoretical framework of learning theory. There are a few other omissions although in the reviewer's opinion they are not so significant as those mentioned above.

The order of presenting the material in the revision is much better and far more "teachable" than was the case in the original edition. The book is arranged into five basic sections. The first section presents a brief orientation and introduces the reader to the eclectic viewpoint which the authors attempt to maintain throughout their book. The second and longest section is given over to the precise description of the various representative kinds of abnormal behavior. The order of presenting the descriptions keeps the various headings such as neurosis and psychosis distinct from each other. In the first edition this was not the case, and jumping back and forth was often quite confusing. The third section presents an extremely fine review of the explanations of abnormal behavior dealing with such varied topics as heredity, culture, development, the brain, etc. That these authors are clearly at home with their material is quite evident from an examination of this section. The fourth section deals with the general features of psychopathology, which are classically treated under the heading of disorders of sensation, emotion, intellect, etc. The fifth and last section, headed "Diagnosis and Therapy," is improperly named since it includes much more than is generally considered under these specific topics. The reviewer feels that too little attention is given to diagnosis and therapy, even though these are usually topics included in so-called "clinical psychology." The chapter on the law as applied to abnormal psychology is especially good and should be carefully studied by all psychologists.

Though the authors state clearly that they intended to approach abnormal psychology from an eclectic and unbiased viewpoint, there is an unmistakable leaning toward biological interpretation. This leaning, in the reviewer's opinion, is easily defended in view of the greater number of precise and objective biologically oriented studies of abnormal behavior than can be found in the social field. Nonetheless, the authors have maintained an exceptional degree of open-mindedness while still preserving the skepticism of the scientist.

The authors state that if they "be allowed a prediction, it is that the next five to ten years will provide more additions to our basic knowledge of mental abnormality than have been made during the past fifty years." It is the reviewer's conviction that should the authors' prediction turn out to be true, they could do no greater service to their profession than to revise their book again.

ROBERT P. FISCHER.

Marietta College.

WATKINS, J. G. *Hypnotherapy of war neuroses.* New York: Ronald, 1949. Pp. x+384. $5.00.

With incisive style, Watkins writes about the skillful use of hypnotherapy with diverse problems. The book is of definite value to those interested in hypnosis, either for research or more pragmatic purposes. Part I is somewhat repetitious of well-known facts about neurosis, but Parts II, III, and IV discuss in detail techniques of induction, palliative therapy with individual cases, and advanced therapy with an entrenched phobia.

Whereas the more nondirectively oriented reader will be shocked by some of the procedures the author employs, the latter is frank about the limitations of hypnotic and directive methods generally. His basic aim in the military setting is not "to rebuild the patient's basic personality structure but . . . to relax, if not resolve, a few of the major unconscious conflicts and to strengthen the ego . . ." This reviewer was particularly impressed by the skill with which various nuances of hypnotic behavior were employed. Among these may be cited the "in-and-out" method wherein the patient is successively brought into and out of trance, progressively recalled memories, automatic writing, and re-adjusted memories. (A critical point, however, is that "adjusted" memory may be ineffective in that it may not provide a completely affective working through. Although this limitation is acknowledged, more research is needed on the length of time the changed memory persists, how well it is integrated, etc., before the method becomes widely practiced as a technique.)

One may take issue with the author's contention that antisocial behavior may be induced under hypnosis, although he does admit the evidence is controversial. R. W. White, for one, seems to favor the view that such behavior cannot be suggested. This is certainly an area where research must be pursued to shed important light on the superego.

A minor criticism is the insertion of a glossary defining obvious terms, which is redundant since the book is geared for the professional. However, this does not detract from a readable and useful casebook.

University of New Mexico. RALPH D. NORMAN.

Vol. 48, No. 4 July 1951

Psychological Bulletin

THE SENSORY CAPACITIES OF INFRAHUMAN MAMMALS: VISION, AUDITION, GUSTATION

PHILIP ASH
The Pennsylvania State College

I. INTRODUCTION

In spite of the voluminous literature on animals and animal experimentation, huge lacunae exist with respect to detailed and reliable information concerning their sensory capacities and limitations. R. L. Yerkes has pointed out this shortcoming with respect to the chimpanzee (**181**) and the other great apes (**185**).

N. L. Munn's 1933 text on animal psychology, which carried the subtitle, "The Behavior of the Rat" (**105**), devoted a scant thirty pages to the sensory processes of the animal that had been the subject of research for over a quarter of a century. In his recent monumental volume (**107**) reviewing psychological resarch on the rat to 1950, much more information concerning the sensory capacities of this animal is available, largely as a result of research since 1930. He points out (**107**, p. 180), however, that although in no other animal, with the possible exception of man, has sensitivity been as thoroughly investigated, large gaps in knowledge still exist with respect to color vision, hearing, and the limits of sensory differentiation in other modalities. The paucity of information available, however, can be appreciated only when the fragmentary facts are assembled and compared with the extent of the animal kingdom. It is the purpose of this review to collect the quantitative data with respect to the limits of visual, auditory, and gustatory capacities of infrahuman mammals, to indicate the coverage in terms of species and experimental populations, and to describe and compare methods employed to determine these capacities. The choice of topics and topic subdivisions has been largely dictated by the volume of data available. Vision and audition both represent areas of considerable work; gustation has been treated in a driblet of papers; and all other receptive capacities combined are relatively terra incognito.

General reviews may be found in the comparative psychology texts, particularly Watson (**167**) to 1914, Washburn (**163**) to 1926, Maier and Schneirla (**99**) to 1935, Warden, Jenkins, and Warner (**160, 161**) to 1936, Fields (**42**) to 1946, and Harlow (**65**). Yerkes (**181**) and Yerkes and Yerkes (**185**) have summarized the data for chimpanzees and the great apes, and Ruch (**134**) has prepared a rather complete (to 1941) bibliography of the literature on primates. Munn (**105, 107**) summarized the data for the rat. Razran and Warden (**124**) reviewed the findings with respect to the sensory capacities of dogs as determined by Russian workers using the conditioned reflex method (up to 1929). Honzig (**73**) and Ghiselli and Henry (**49**) briefly covered the role of sensory acuity and the sensory thresholds of rats in relation to maze learning. Annual reviews of current findings relating to sensory processes are given by Davis (**29**), Hartline (**67**), and Olmstead (**112, 113**). These reviews contain brief references to studies of mammalian sensory capacities.

There seems to be little question concerning the psychological importance of investigations of the sensory capacities of numerous mammalian species. The animal is well established as a research tool to supplement work on man and to provide information that cannot be obtained with human subjects. It should be stressed, however, that no transfer from infrahuman to human behavior can be made with confidence until contingent similarities and differences in sensory capacity are determined and evaluated. The selection of animals for particular studies, the development of appropriate apparatus, and the formulation of suitable experimental designs depend, or perhaps *should* depend, upon the characteristics of the animals themselves. Fundamental among these characteristics are the limitations and response ranges of the sensory modalities through which stimuli are received. Unless the limits of these sensory modalities are known for large groups of animals, intelligent selection of experimental populations and adequate experimental designs are difficult or impossible.

II. Vision

More has been written about vision in mammals than about all the other sensory modalities combined.

The field of vision has traditionally been broken down into a variety of subfields: visual acuity, brightness discrimination, color vision, movement discrimination, flicker discrimination, depth discrimination, form and pattern discrimination, size discrimination, and distance discrimination. In the majority of these areas, however, substantial data concerning threshold functions are lacking. For example, various investigators have been able to demonstrate brightness discrimination behavior in mice, rats, guinea pigs, rabbits, chickens, raccoons, porcupines, cattle, dogs, monkeys, cats, and horses. But in spite of the fact that this is one of the easiest discriminations to establish, variations in apparatus and methods have made it almost impossible to state comparable threshold values. Form discrimination has also thus far not yielded to adequate standardization which would permit cross-species comparisons. This paper will therefore limit consideration to studies dealing with the measurement of visual acuity thresholds and color vision, with incidental reference to the role of brightness in color response behavior.

In addition to the general treatments referred to in the introductory paragraphs, vision in mammals (anatomy, physiology, and visual capacities) is treated in Bartley (**8**, **8a**), Detwiler (**30**), Granit (**53a**), Parsons (color vision) (**115**), Rochon-Duvigneaud (**131**), Stagner (**145**), and Weymouth (**179**). Detwiler's *Vertebrate Photoreceptors* and Bartley's *Vision* are especially valuable for their accounts of the visual mechanism. However, the data on the response capacities of mammalian species are conspicuous by their absence. The visual literature has been catalogued by Fulton, Hoff, and Perkins (**44**) in a bibliography of over 3,400 items.

Visual Acuity

Redetermination of the visual threshold of man by Hecht, Schlaer, and Pirenne under the "most favorable conditions of dark adaptation, peripheral vision, small test fields, short exposure, and selected portions of the retina" yielded an absolute threshold of 2.1 to 5.7×10^{-10} ergs in the blue-green, corresponding to between 54 and 148 quanta at the cornea, and probably 5 to 14 quanta absorbed by the rods, or an absorption of about 1 quantum each by 5 to 14 rods out of 500 in the geometrical image of the stimulus. With so few quanta involved, human thresholds should (and do) show great variation. It is to be anticipated that similar variation, based upon the same considerations, should appear in other species.

Minimum separable and minimum visible thresholds. Minimum separable acuity (the acuity of distinguishing two lines in close apposition) and minimum visible acuity (the acuity of distinguishing a single line against a homogeneous background) have been regarded as dependent upon the same processes, but the thresholds for the two functions vary greatly. In man, minimum separable acuity varies between 1.0 and 0.5 minutes of arc, while minimum visible acuity approximates 0.066 minute. With the exception of Klüver (**90**), Warkentin (**162**), and

eau Ednl. Psy. Research
AINING COLLEGE

Smith (**142**), however, determinations in animals have been in terms of minimum separable acuity only (see Table I).

Methods. The discrimination technique has constituted the primary method employed in the study of visual acuity in animals. Some modification of the Yerkes-Watson discrimination box (**184**) has been used by Spence (**144**) and Johnson (**84, 85, 86, 87**). Lashley employed the jumping technique he devised for the study of pattern vision (**95, 96**). Smith (**141, 142**) employed a combined discrimination and conditioning technique, which seems to be a version of the Yerkes-Watson apparatus in which an electric shock punishment is introduced before the animal reaches a goal point. A stimulus box is set at the end of each of two runways. After release from the cage, the animal is required to make a discrimination response at the end of the runway table away from the boxes. If the response is correct, the animal may approach the box and obtain the food by pressing a lever. If the response is incorrect, it is shocked at the beginning of the runway.

In the discrimination studies, the stimuli characteristically are modified Ives instruments which present the animal with visible striae. One test field is adjusted to a point where it appears homogeneous, and in the other the striae are gradually reduced until a discrimination between a striped and homogeneous field breaks down. Another method employed involves setting up the fields so that the striae in one field are at right angles to the striae in the other. The separations are then reduced until this discrimination breaks down. Johnson (**85**) found that the two methods yielded comparable results.

Warkentin (**162**) determined visual acuity thresholds by means of oculocephalogyric responses to a moving visual field. The animal was placed in a rotating cylinder that contained vertical black and white striations which took the place of the Ives gratings. Smith and Warkentin (**143**) employed optokinetic responses in a similar manner.

Klüver (**90**, p. 227 ff.) employed an unlearned response (grasping) technique in his work with monkeys. The animal is confronted with a box near the top of which is an opal glass, illuminated from the bottom. Within reach of the animal, and lying on the glass, is a thread to which food is attached. If the thread is visible, the animal grasps it and pulls in the food. The experiment is performed in a dark room with illumination available only from behind the opal glass. Klüver also used the "pulling-in" technique in a variation of the traditional discrimination experiment. The animal faced two or more boxes, above one of which was placed a positive stimulus field (e.g., a cardboard with a line drawn on it) while above the others a blank card was placed. The animal was trained to pull toward him the box under the positive stimulus field.

While the results obtained from this last method may provide a rough index of acuity or discriminative capacity, the method as used is probably too uncontrolled in terms of brightness, accommodation distance, and especially measurement to yield very exact threshold values.

Results of visual acuity studies. Visual acuity has been demonstrated to be a function of the test object, accommodation distance, general field illumination, and illumination of remote regions (**8**, p. 34 ff.), when stimulus characteristics are considered. These are quite probably the minimum variables from the stimulus side. Yet the reported animal studies do not specify all these factors, nor are sufficient data available to reduce to a common denominator acuity measurements resulting from different test situations. Table I presents visual acuity figures found in the literature for mammalian species.

It is to be noted that wide interspecies differences manifest themselves. However, in view of the extremely limited character of the samples employed, it may be that intraspecies variation is still greater, except for the most divergent species.

Visual acuity and other visual functions. Visual acuity, as distinguished from brightness, size, distance, and form discrimination, shows little relationship with phylogenetic status (**162**), although from a structural standpoint great interspecies differences may be noted with respect to the fineness and distance apart of the receiving elements and in the number of conducting and association elements. It is therefore possible that phylogenetic differences with respect to discriminations other than acuity are to be accounted for, not in terms of any improvement on the optical efficiency of the eye, but in perceptual processes associated with vision and in the intelligence of the animals in solving the problems by means of which an experimenter measures brightness (or other) discrimination. On the other hand, precisely these interspecies structural differences may be invoked to account for visual acuity differences. Grether (**63**) reviewed the data on minimum separable acuity thresholds for the chimpanzee, rhesus monkey, Cebus monkey, rat, pigeon, and gamecock, and used the data for a computation in terms of retinal image width. He concluded that thresholds thus expressed provide a more direct basis for interspecies comparison of ocular efficiency than do visual angles. Because of their smaller eyes, all animals considered compared favorably with man when retinal widths were substituted for visual angles as a measure of acuity.

TABLE I

VISUAL ACUITY THRESHOLDS IN INFRAHUMAN MAMMALS*

Animal	*N*	*Method*	*Accommodation Distance* (cm.)	*Brightness* (mll.)	*Minimum Separable Acuity* (min.)	*Minimum Visible Acuity* (min.)
Rhesus monkey (171)	2	Discrimination box	72	37.0 0.75	0.90 0.67	
Chimpanzee (144)	2	Discrimination box	120	3.98	0.47	
Cebus monkey (85)		Discrimination box	60	3.85	0.95	
Java monkey (90)	1	Thread-grasping ("pulling-in" technique)	28	(0.007)†		2.5
Cebus monkey (90)	1	Discrimination—single thread ("pulling-in" technique)	200	Sunlight		0.044
Albino rat (96)	3	Jumping technique	20		52–86	
Pigmented rat (96)	5	Jumping technique	20		26–52	
Cat (142)	2	Discrimination and conditioning	50–120		5.5‡	0.45–3.3
Domestic rabbit (162)		Oculocephalogyric response				1.4–1.8
Wild rabbit (162)		Oculocephalogyric response				0.50
Guinea pig (162)		Oculocephalogyric response				.50–1.4

* All studies but Warkentin's (162) used visual striae produced by some version of the Ives apparatus.

† Foot-candle illumination.

‡ Not a threshold, but the minimum obtained with the apparatus.

The precise meaning of this conclusion for comparative studies is not clear, however. A four-cylinder gasoline engine will yield more miles per gallon than an eight-cylinder engine. If a comparison is to be based upon gasoline consumption, the smaller engine is "better." However, if other performance characteristics such as acceleration, speed, and hill-climbing ability are taken as criteria, the larger engine invariably proves the "better." If interspecies comparisons are to be made in terms of the dimensional characteristics of the "real world," then the latter kind of analysis is assumed. An interesting experimental problem is posed, however: What is the relationship between acuity of sensory discrimination and adequacy of perceptual and cognitive organization? Ghiselli and Henry (**49**) found a correlation of .31 between differential proprioceptive sensory thresholds and learning in rats, and one of .36 between sound thresholds and learning, when the stimuli were above the absolue threshold. Too frequently, however, the absolute threshold is not known or considered. For example, if the rat's eye is as efficient as the human eye when the physical stimulus components are equated in terms of comparable retinal image width, then under these conditions what are the characteristics of rat learning? It is suggested that learning experiments might be arranged to take these factors into account. Typical apparatus design calls for scaling equipment down to the size of the animal, whereas these considerations suggest that the size of the test object must be increased as retinal image width decreases if the level of difficulty from point of view of basic sensory discrimination is to be held constant.

It seems clear that the interrelationships of these factors have not yet been adequately explored, and that the data essential to undertaking such investigations are still lacking. In this connection, Harlow (**65**) points out that no discussion of visual discrimination is complete without noting the great emphasis placed on utilizing efficient apparatus, in contrast to the comparatively slight emphasis placed on utilizing efficient stimuli. Kohts (**183**), for example, found that stereometric objects are more easily discriminated by the chimpanzee than are the types of stimuli (designs painted on flat surfaces) commonly used in animal laboratories, and it is not unlikely that this finding would hold good for many other species.

Color Vision

One of the most controversial topics in animal psychology has been that of color vision in species below man. In the first place early experimental evidence suggested it was lacking in such favorite laboratory animals as the rat. In the second place, theories of vision that consider color discrimination dependent upon an evolutionary process in the eye (and color blindness in man an evidence of atavism) seem to require that color vision, if it exists below the primate level, be very rudimentary at best (**58**). The doctrine that color vision was probably lacking below the primate level in mammals therefore became an accepted part of the literature, and led Munn to conclude that " . . . the rat is color-blind . . . [and] these results . . . are amply substantiated by many other experiments involving infra-primate mammals, all of which appear to lack cones" (**105**, p. 131).

The more recent evidence has reversed or substantially qualified this conclusion, although Munn, in his latest summary of the issue (**107**, p. 143), suggests that the existence of color discrimination in the rat is still so controversial a question that further research to refute or support the available positive results is required.

Three lines of evidence are available for the study of mammalian color vision: first, experiments in which the animal is trained to make a positive response to one color and no response or a negative response to another; second, anatomical and histological studies of the eye to determine the presence or absence of cones; and third, electroretinographic studies to determine by comparison of the spectral curve in the light-adapted eye with the curve of dark adaptation, whether the differences characteristic at the human level appear (Purkinje phenomenon, displacement of photopic curve).

This paper will be limited to the first line of evidence primarily, with reference to a few of the findings concerning the spectral curves of some species.

Methods. A wide variety of techniques have been employed to determine the existence of color vision in mammals. The most frequently used apparatus has been the Yerkes-Watson discrimination box or some modification of it (**21, 89, 103, 104, 106, 108, 155, 157, 164, 168**). Walton and Bornemeier (**158**) and Russell and Walton (**135**) worked with the Lashley jumping apparatus. The conditioned response technique was used in Pavlov's laboratory (**120**), by Walker in Bekterev's laboratory (**124**), and by Brown (**18**). Klüver had monkeys pull in strings to which differently colored boxes were attached (**90**).

As wide a range of light sources as techniques have been utilized. The most precise light source, spectral colors, was used early by Watson (**166, 168**) and more recently by others (**18, 53, 59, 75, 149**). More frequently, however, monochromatic light was obtained by means of Wratten or other filters (**103, 106, 108, 155, 157, 158**). Hering and

Bradley papers (**10**, **75**, **89**, **104**, **164**, **169**), dyes (**139**), paints (**21**, **22**), colored disks (**55**), cards (**19**), and the color wheel (**92**) also have provided the color stimuli.

This variety in method has without doubt contributed to the lack, or apparent lack, of consistency in results, particularly in the case of animals with which a large number of experiments have been performed. The most serious shortcoming, one which seems to be common to almost all the earlier experiments, is the absence of adequate control of the brightness factor (**145**).

In the first place, adequate brightness control is difficult or impossible when papers or paints are used, and presents some problem with filtered lights that are not strictly monochromatic. In the second place, in some instances brightness of the stimulus colors was matched for the human eye. Infraprimate mammals generally react much more readily to brightness differences than to hue differences, and it has been demonstrated that hues matched for brightness to the human eye may appear of different brightness to the animal's eye. Current technique therefore calls for the selection of a pair of hues demonstrably equivalent in brightness for the experimental animal, and subsequent training of different animals from those on such a pair of colors used for the selection ("confusion colors"). Finally, after differential discrimination to this pair is established, control experiments in which the brightness relations of the colors are systematically varied are necessary.

Experimental evidence on the presence or absence of color vision is available for only a very few species, and conclusions in which any confidence can be placed are possible only for primates, rats, and possibly dogs.

Squirrel and porcupine. Sackett (**136**) reported positive findings with respect to the existence of color discrimination in the porcupine, and Colvin and Burford (**22**) claimed evidence that the squirrel could discriminate colors, but in neither case was brightness controlled.

Raccoon. Cole (**19**) and Cole and Long (**20**) reported positive results, but brightness was not controlled, or it was merely equated for the human eye. Davis (**28**) found that his animals were responding to a brightness difference, rather than a hue difference. Gregg and McPheeters (**56**) obtained some evidence of color discrimination, but in a later study Gregg, Jamison, Wilkie, and Radinsky (**55**) found that the animal could not discriminate pairs of colored disks.

Calf. Kittredge (**89**) trained a calf to discriminate red from a Hering No. 5 gray, but the animal could not discriminate the red from a No. 50, suggesting that it was responding to a brightness difference.

Dog. Colvin and Burford (**22**) trained three dogs to pick a positive color out of nine colors, but brightness and experimental cues were not controlled. Samojleff and Pheophilaktowa (**137**) found that their dog could discriminate green from most grays, but that one or two of them were frequently confused. However, even these grays were discriminated from the green only after extended training. Smith (**140**), using transmitted light, found that some dogs apparently could make color discriminations, but he concluded that the discrimination was very unstable "and cannot be supposed to play any part in the animal's normal existence." Walker, in Bekterev's laboratory, was able to establish a conditioned color response only after 3,000 trials (**124**). Pavlov found that the salivary reflex could not be conditioned to a color stimulus (**120**, p. 101). Yerkes and Morgulis (**182**) similarly reported that Orbeli failed to establish a conditioned respose to color in several hundred trials. They concluded that "since this is the most delicate method which has probably ever been devised for studying differential sensitivity, it is not likely that the dog has color vision."

Guinea pig. Allen (**3**), employing a Yerkes apparatus, found no evidence of color discrimination in this animal.

Cat. Devoss and Ganson (**31**) found that cats could discriminate colors from gray papers, but not from gray cambric. Colvin and Burford (**22**) trained a cat to pick a positive color out of a group of nine colors, but brightness was not controlled; Gregg, Jamison, Wilkie, and Radinsky (**55**) failed to establish discrimination between three pairs of colors. Granit (**54**) determined the curve of photopic vision by placing electrodes on the retina. He found the point of maximum sensitivity at 560μ, and a spectral curve that approximated the human curve of photopic vision. This finding does not seem to be decisive, however, because Graham, Kemp, and Riggs (**52**) were unable to establish a cone vision curve in the pigeon, which does discriminate colors.

Rabbit. Washburn and Abbott (**164**) found that the rabbit could not discriminate between two doors leading to food when the stimuli to be differentiated were a red and a Hering No. 46 gray. However, Brown (**18**), using conditioned respiratory responses, determined the brightness visibility curve and the dim visibility curve of the rabbit. He found that the dim visibility curve was identical with the human dim visibility curve and the brightness curve was displaced 30μ toward the red end of the spectrum. Further, he demonstrated a consistent differentiation between red and green in spite of systematic variations in brightness.

Rat. The bulk of work on color vision below the primate level has been done with rats (and the writer will include mice). At least until 1933 the evidence has been interpreted to mean that these animals lacked color vision (**21, 74, 75, 76, 97, 104, 106, 168, 169, 180**). Munn's (**105**) earlier comment in this connection has been cited above.

The anatomical evidence is not so clear-cut, however. Vincent (**151**) and Lashley (**97**) found no evidence of cones in the rat's eye, nor Hopkins (**74, 75**) in mice. However, Detwiler (**30**, pp. 22–23) reported the measurement of the cone layer in the eyes of various vertebrates. He listed positive measurements for the marmoset, rhesus monkey, and chimpanzee; indicated that accurate measurements were uncertain in the dog and cat because of pigment masking or unsatisfactory histological preparations; listed "very sparse or absent" for the rat and hamster; and stated that cones were lacking in the bat, lemur, and owl monkey. Walls (**153, 154**) found that rats *do* possess cones, but he is still skeptical of the results of recent studies indicating that the rat can make at least limited color discriminations. To complicate matters further, Graham and Riggs (**53**) recorded the retinal potentials from the eye of the rat in response to lights of various wave lengths. They found that the "retinal response of the rat's eye is one which varies only in relation to brightness; a response for any wave-length may be matched by a response for another wave-length provided the proper intensity is found for the second wave-length." Then, in a later study, Graham, Kemp, and Riggs (**52**) were unable to obtain a cone curve from the eye of the pigeon, which has color vision.

The history of color vision experiments with rats is given in detail in Munn (**107**, pp. 134–143). By 1933 the accumulated evidence suggested that the animal was color-blind. Then color discrimination was demonstrated by Walton (**155**); this study was criticized for inadequate control of brightness by Munn (**106**), who seemed to demonstrate that Walton's findings might in fact have been due to the brightness variable rather than the hue variable. Then, Muenzinger and Reynolds (**103**) obtained a discrimination on a discrimination box apparatus between red and green, and one between red and black. Their results were confirmed by Munn and Collins (**108**). Walton (**156**) and Walton and Bornemeier (**157, 158**) extended and perfected their work, using a discrimination box and a Lashley jumping apparatus. Discrimination between matched pairs of yellow-blue, red-green, and red-blue, somewhat unstable discrimination between red and yellow, and no discrimination between blue and green or between yellow and green were found. It was significant that while a brightness discrimination could be established in an average of 85 trials, a hue discrimination required between 375 and 525 trials. In some cases Walton and Bornemeier employed 800 trials to obtain positive results.

Most recently, Russell and Walton (**135**) demonstrated a red-blue discrimination in 18 rats, albino and hooded, subjected to avitaminosis-A. As the avitaminosis continued, accuracy of response was maintained at 95 to 99 per cent. These results indicate that avitaminosis-A, which has a marked effect on rod vision, may have no effect on cone vision. Munn (**107**, p. 142) suggests further that the failure of avitaminosis-A to disturb the red-blue discrimination may support the conclusion that the discrimination is based on hue rather than on small brightness differences.

Muenzinger and Reynolds concluded that rats possess a hierarchy of discrimination habits consisting of a "primary brightness relation habit, a secondary absolute brightness habit, and a weak color discrimination habit." The relative brightness values of the colors have been established as follows: blue has the highest value for the rat, green and yellow come next, and red has the lowest value.

Infrahuman primates. Positive findings with respect to the existence of color vision were noted in all but one out of about a score of papers. Color vision has been demonstrated in the chimpanzee (45, 59, 60, 61, 62, 64, 91, 92, 183), the orangutan (45), and in the Java, Cebus, spider, and rhesus monkeys (10, 58, 88, 90, 139, 147, 166). It has been demonstrated with papers, the color wheel, painted disks, and spectral colors. It has repeatedly been shown that such discriminations are independent of brightness (10, 58, 59, 60, 61, 64, 139, 183).

With this array of evidence available, it becomes possible to review the specific characteristics of this capacity in infrahuman primates. Early research (10, 88) suggested that color discrimination is easier for these animals than brightness discrimination (discrimination among grays). However, this result may have been due to interference with previous training rather than an inherent limitation. For example, Bierens de Haan (10) presented his subjects (two pig-tailed macaques) with a five-choice apparatus. Each of five boxes had a color on one side: white, black, red, yellow, or blue. The animal was trained to respond to, say, the white. The boxes were systematically shifted about so that the white was distributed according to a chance sequence throughout the trials. After discrimination of the white from the other colors had been established, the other colors were replaced with grays, and the discrimination habit was found to persist. The same procedure was employed with each of the five colors as the positive stimulus. However, when a gray was made the stimulus color in a series of five grays, no discrimination habit could be formed. In addition to the possibility here that brightness was not adequately controlled, it seems likely that the second habit interfered with the first. Kuroda (93) repeated Bierens de Haan's experiment and obtained negative results, the only study the author has noted in which this is true. Kohts (183) trained a chimpanzee to select, from forty colored disks, one to match in hue a disk offered by the experimenter. She noted, however, that achromatic stimuli were not as well distinguished as chromatic stimuli.

The most intensive work with primates has been undertaken by Grether. His careful explorations of the limits and characteristics of primate color vision are models of experimental investigation. He used monochromatic lights with a discrimination apparatus in which the animal was required to take food out of a box on which the positive stimulus color was projected. In one experiment (58) with monkeys he determined the limens for hue discrimination as follows: the animal was trained to respond positively to yellow (589μ) and negatively to blue-green

(500μ). Training continued through changes in brightness relations until the color habit was fixed. Then the wave length of the negative color was increased 20μ, and in consecutive trials increased gradually until discrimination between the yellow and the negative color broke down. A criterion of 75 per cent correct in 25 trials was set, and each wave-length step was offered in 25 trials. The procedure was repeated to obtain limens for red and blue-green. Table II presents the limen values in μ for the animals tested. Except for the Cebus monkeys' response to the red and yellow, these limens are almost identical with the human limens.

TABLE II

Color Vision Limens in Monkeys in Mu
(From Grether [**58**])

Primate	*Stimulus Color*		
	Red (640μ)	*Yellow* (589μ)	*Blue-green* (500μ)
Cebus Monkey	30–45	4–8	7–10
Spider Monkey	6	2	10
Old World Monkeys	8–12	2–4	7–10
Man	10–12	1–3	6–21

Repetition of the experiment with chimpanzees (**59**) showed that chimpanzee limens for hue discrimination are approximately equal to that of man in the blue-green, but about twice that of normal human vision in the yellow and red. When response of chimpanzees in terms of color mixture proportions was compared to man's, it was found that chimpanzees required more red in a red–blue-green complementary mixture to match white, but that the proportions of red and green to make a yellow were the same for the two species. Using a color against darkness under conditions of light adaptation (15 seconds' exposure to a light before making the discrimination), Grether found the spectral limits of chimpanzee vision to be between 402μ and 704μ; for human beings the results were between 399μ and 692μ (**61**). The spectral saturation curves for chimpanzee color vision were determined by finding the proportion of spectral color mixed with white needed to produce a discriminable hue (**64**). The curves for two chimpanzees were very close to those for two human beings, the lowest saturation point for human subjects being at 575μ, and for chimpanzees 570μ (**62**). Grether summarizes his findings with the conclusion that the color vision of the chimpanzee and the rhesus monkey is at an evolutionary position just short of human vision, and that the slightly deficient hue discrimination of these animals could be accounted for by less complete differentiation of the red and green receptor mechanisms. The markedly anomalous hue discrimination of the Cebus monkey is taken as evidence of two-color vision in that animal (**59**). The observation that the crossing point in red-green discrimination (the discrimination habit reverses as the difference becomes subliminal) is about 5μ nearer the shorter end of the spectrum in chimpanzee vision than in normal human vision, is taken to mean that chimpanzee vision slightly resembles red-green color blindness, another evidence of lesser development (**64**).

Grether's findings are highly provocative, not only for the information they provide about primate color vision but for the implications for color vision theory. It seems clear, however, that further corroboration is necessary, since his animal samples are very small (involving at most nine monkeys in one study). Two sorts of questions are pertinent: first, are the differences between infraprimate and human vision noted here within the sampling error, and within a broader range of variation among the animals? Second, to what extent has the laboratory environment, with its limitations and restrictions, interfered with the development of color vision habits in these animals? Are there differences between laboratory animals and animals in their natural setting in this regard?

Color vision may be regarded as an established fact in some primate species. In view of the evidence that owl monkeys lack cones (**30**, p. 23), this species may not be able to discriminate color. A beginning has been made in the determination of the limits and quality of primate color vision. The positive evidence in favor of color vision in rats seems clear, but it has not yet been reconciled in any adequate fashion with the electroretinographic findings. These perhaps may be reinterpreted or revised, or an essential independence between retinal response to spectral lights and color vision may obtain. It seems quite likely that, in a qualified way, the latter alternative will prove to be the more appropriate.

With respect to other mammals, the volume and character of the evidence fail to justify any positive conclusions. The history of experimentation with color vision in the rat should discourage ready acceptance of negative findings with other animals on the one hand, and on the other the failures in early experimentation to control contingent variables, particularly brightness, inspire little confidence in what meager positive findings have been made. Adequate exploration of the color sensitivity of mammals is still a task for the future.

III. AUDITION

Two interrelated experimental problems have been attacked with reference to the modality of hearing: First, can animals discriminate pure tones from noise, and one tone from another? Second, if they can discriminate tones, what are the characteristics of their audiograms?

It is not improbable that mammals differ considerably with regard to the capacity to discriminate sounds on the basis of wave-length differences (**161**, p. 293). The first problem, however, seems to have developed *as a problem* as much through an overzealous application of the law of parsimony as from the experimental data. Its history is reminiscent of the history of the color vision issue. Firm opinions were formed with respect to the guinea pig, the dog, the cat, and particularly the rat, to the general effect that these animals could not discriminate pure tones, or could discriminate them only poorly. It was not denied that they could hear, or that they could discriminate noise from silence. Until the mass of evidence was clear, however, it was assumed that all pure tones sounded alike, or that they were not adequate stimuli for the animals' auditory receptors. With the resolution of this problem in the conclusion that these animals *do* discriminate pure tones, attention was directed toward the second question. Two aspects of it will be explored in particular: the extent of the auditory frequency spectra in infrahuman mammals, and the minimal intensity values (relative to human thresholds) required to evoke a response at various frequencies. To a greater extent than in other sense modalities, the more recent results have been both clear and consistent for several infrahuman mammalian species.

Current summaries of the status and most recent findings with respect to audition have been written by Bárány (**6**), Newman (**111**) and Walzl (**159**). In 1950 the Psycho-Acoustic Laboratory of Harvard University published a bibliography on audition including over fifty-five hundred entries (**190**).

Methods

Discrimination mazes. For the rat, and in a few instances for other animals, the principal early method of determining tonal sensitivities involved some variant of a discrimination problem, in the form of a maze requiring choice between a tone-filled alley and a silent alley, or between two alleys with a relatively great difference in tone intensity (**68, 69, 70, 71, 72, 78, 79, 80, 81, 83a, 102, 148**). The more successful apparatus incorporated relatively intense electric shock. Henry (**69**) found that when a shock of 135 volts was given, discrimination habits were established in from 14 to 70 trials, while a shock of 40 volts had little effect. This may be contrasted with the experience of Thuma (**148**) on a slightly different maze, on which discrimination took between 710 and 1,410 trials, and with that of Muenzinger and Gentry (**102**), in which discrimination required between 80 and 250 trials. In the typical experiment the animal is required to go down the tone-filled alley to the food. A sound source is placed at the end of each alley; the apparatus usually provides for silent switching-on of the sound to a predetermined intensity level. A serious objection to the two-alley maze arises from the fact that even the best sound-conditioning permits some sound in the "silent" alley. For example, Henry and Brown (**71**) found that while the active alley of their apparatus had an intensity level of 50 decibels at 2,500 cycles, the inactive alley had an intensity level of 38 decibels.

Conditioned response techniques. In the more recent studies with rats, and in the great majority of other studies, the conditioned response technique has been the preferred method. Britt (**15, 16**) and Cowles and Pennington (**24**) used "classical" conditioning, conditioning the rat's respiratory response and squeak, respectively, to tones. Conditioned shock-respiratory response has also been employed with cats (**173**), and guinea pigs (**77, 150**). Black and Schlosberg (**12**), Eccher (**38**), Gentry (**48**), Gould and Morgan (**50, 51**), and Hunter and Pennington (**82, 83**) established an "instrumental" conditioning with the rat, involving training it to move from one part of the apparatus to the other to escape shock. Dworkin and his associates employed a conditioned motor-alimentary response with cats and dogs: the animal was trained to obtain food in response to the stimulus tone by raising the hinged lid on the food box (**33, 34, 35, 36**). The method is described by Dworkin, Sutherland, and Seymour (**37**). Ades, Culler, Mettler, and their associates utilized a motor flexion response (**1, 2, 17, 26, 27**). The salivary reflex has been used in Pavlov's laboratory and by others (**4, 5, 120, 124**). Sutherland and Dworkin (**146**) compared results from the salivary and motor-flexion responses, and Dworkin (**35**) compared the salivary response with the alimentary-motor response. Elder (**39, 40**) conditioned chimpanzees to press a telegraph key at the sound of a tone to release food from a hopper. Harris (**66**) and Wendt (**172**) used conditioning techniques with monkeys to determine their tonal sensitivities.

The conditioned response technique, as applied to studies of tonal discrimination, has the following general characteristics: when the problem involves determination of upper frequency limits, conditioning is effected at a supraliminal frequency, and the frequency is increased until the response is no longer made. Occasional reinforcement (substitute reinforcement) at liminal values may be necessary to obviate the possibility of extinction. When the problem involves determination of an audiogram, training is effected at a supraliminal intensity at each frequency measured, and then the intensity is systematically decreased. To determine tonal differentiation, a positive response is conditioned to one tone and a negative response (inhibition of response) to another which is supraliminally removed in the auditory spectrum. Then the negative tone is gradually reduced (or increased) in frequency until the differential response fails. Since failure of differentiation is usually accompanied in each type of experiment by symptoms other than mere extinction, such as tantrums, increased latency of response (**12**, **66**), negativism (**5**, **39**, **40**), marked behavior disturbances (**36**, **39**, **40**), and the like, the determination of liminal values can be rather clearly effected.

Methodological problems. The conditioned reflex method of determining animal auditory spectra has been, in general, the most satisfactory and has given the most consistent results (**39**). It provides the closest approximation to the audiogram that can be obtained in experimental animals (**159**). Thresholds obtained from it have been lower than those from mazes (cf. **45**, and **12**, **16**, **24**, **38**, **51**). However, interpretation of these audiograms, especially in terms of comparison with human thresholds, is somewhat beclouded by two factors.

First, the available evidence shows that practice lowers the threshold and would therefore tend to provide estimates of the sensory capacity of the experimental animal that are at variance with its typical, i.e., unpracticed, capacity. Drury (**32**) found that trained telephone operators showed a 12.7-decibel advantage over untrained operators, and Elder (**39**) noted a 5- to 10-decibel improvement in two adult "controls" he employed in his study of chimpanzees. In animal conditioning experiments a considerable practice effect is possible. Further, the amount of practice (trials required for establishing a discriminating response) has varied with the animal, the technique, and particularly the frequency. Dworkin, Katzman, Hutchinson, and McCabe (**36**) reported that it required three to four months of training cats and dogs in the alimentary-motor response before testing could be done, while Lipman and Grassi (**98**) required only two to four weeks with dogs. Razran and Warden (**124**) reported that, in the Russian laboratories, some discriminations in the dog were effected in 10 to 100 trials, while others required between 100 and 1,000 trials. Hunter and Pennington (**82**) reported that it took 325 trials to train a rat to respond to a 1,000-cycle tone (and only 70 trials to train the animal to respond to a buzzer). Since, except in the audiometric laboratory, the animal has generally had no intensive practice in tonal discriminations, it would be desirable for many purposes to partial out the effect of practice on its audiogram and thus arrive at a "true" (i.e., typical) estimate of its auditory capacity.

Second, the observation of a number of workers that the auditory cerebral cortex is not essential for conditioning to a sound stimulus[1] (**123**, **132**, **159**) makes it important to know which auditory functions can be subserved by each level of the higher brain pathway if conditioning is to be used as a testing method. Further, Neet's (**109**) finding that, with human subjects, verbal report yielded lower thresholds than either a conditioned respiratory response or a manual motor response adds another caution to comparisons between animal audiograms based on conditioning experiments and human audiograms based on report.

---

[1] Sutherland and Dworkin (**146**), however, found that after bilateral surgical lesions in the auditory cortex of the dog, no clear-cut CR could be elicited by pure tones.

Sound sources. Early experimenters used sound sources of variable acoustic properties, such as tuning forks, reeds, vibrators. With the development of electronically driven audio oscillators, tones of high purity may be readily generated (25, 71). Furthermore, Dworkin's (36) objection to the effect that the characteristics of sound transmitters at high frequencies are not known is probably no longer valid. Particularly since World War II audio instrumentation and technology have made rapid progress (111).

Auditory Capacities of Selected Species

Audiograms have been determined under relatively comparable conditions, for infrahuman primates, dogs, cats, guinea pigs, and rats. The evidence indicates that the upper limit of the auditory spectrum of all these species (with the exception of the guinea pig and possibly the cat) is higher than man's, and that the intensity threshold is lower at the higher frequencies. Figure 1 presents typical audiograms for several species.[2]

Primates. The auditory capacities of monkeys were investigated by Harris (66) and Wendt (172). Intensity limens were calculated from 62.5 cycles to 8,000 cycles (66), and from 64 cycles to 16,384 cycles (172). Harris employed the psychophysical method of serial groups, while Wendt determined limens by setting the threshold at the lowest intensity to which the animal responded (omitting an ascending series). The results were highly consistent, however, particularly with respect to the shape of the audiogram. The monkey's thresholds were slightly higher than man's in the middle range (1,000–2,000 cycles) and considerably higher (9–25 decibels) at 4,000 cycles, where a marked "dip" was noted. The animals' threshold was lower (14–30 decibels) at 8,000 cycles and above.

Elder (40) placed the upper limit of the chimpanzee's auditory spectrum at between 26 kilocycles and 33 kilocycles (as compared with 22.6–23.7 kilocycles for three children), and determined an audiogram (39) essentially similar to that for monkeys, including the "4,000-cycle dip."

---

[2] These audiograms are approximate. In some cases, values were estimated from figures in the sources cited. Furthermore, the reference level presented here—average human sensitivity—is not constant for all studies cited. The audiogram for the rat is an average of those given by Black and Schlosberg (12), Cowles and Pennington (24), Eccher (38), and Gould and Morgan (51). Henry's (70) audiogram was too discrepant from these to warrant including in the average.

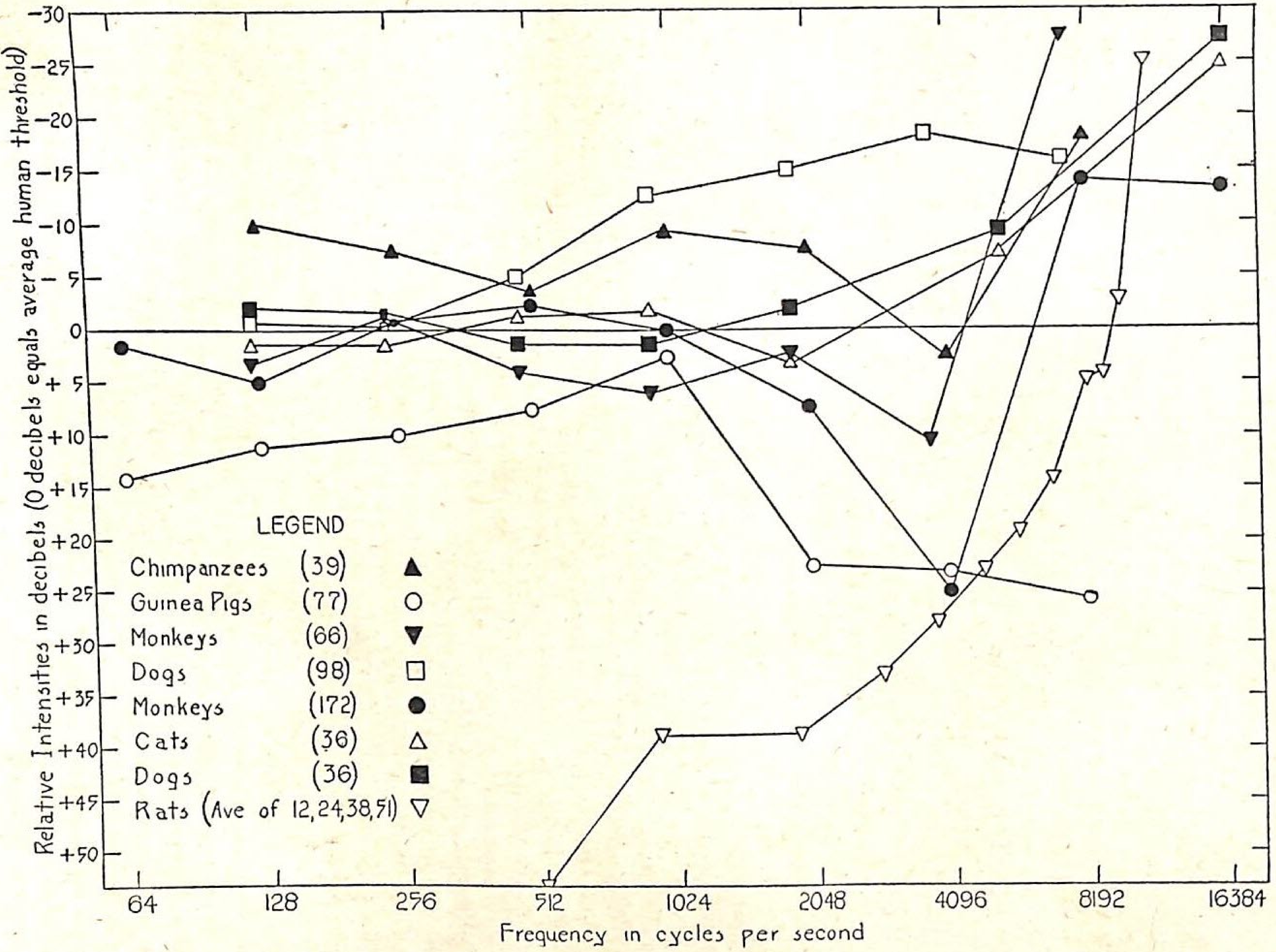

FIG. 1. COMPARATIVE AUDITORY ACUITY THRESHOLDS OF SEVERAL MAMMALS.

Cats. Using conditioning methods, the audiograms of cats were determined from about 100 cycles to 16,000 cycles (**1**, **2**, **33**, **36**, **109a**). Conditioned responses were elicited at frequencies as low as 30 cycles (**33**). The limens were lowest between 5,000 cycles and 8,000 cycles, and the animal's threshold was appreciably lower than man's at frequencies above 5,000 cycles (**36**). At 14 kilocycles to 16 kilocycles the animal's hearing was about 25 decibels better than man's. Wever (**173**) determined the upper limit of frequency response to be between 15 kilocycles and 20 kilocycles.

Some study has been made of the cat's ability to discriminate differences in tone. Dworkin (**34**) trained cats to make a conditioned motor-alimentary response discrimination between 2,150 cycles and 2,360 cycles, or a difference of about one tone. He also explored the cat's ability to make intensity discriminations and found for each of two frequencies that the limit of intensity discrimination was relative rather than absolute. That is, if the positive intensity is 54 decibels and the negative 50 decibels, 54 decibels will be responded to as the negative if it is preceded by the tone at 58 or 60 decibels. He noted that whenever pitch discrimination was developed in the presence of a notable intensity difference, the significance of the notes could be reversed by reversing the intensity. This finding is reminiscent of the relations obtaining between brightness and hue discrimination in the rat.

Dogs. Criticizing the earlier work of Pavlov's laboratory that suggested that the dog had a high degree of pitch discrimination, Johnson (**83a**) employed a Yerkes-Watson discrimination apparatus in an attempt to train dogs to discriminate between a "pure" tone and silence, and between two pure tones. He found no evidence of learning, and came to the conclusion that the animal's capacity for pitch discrimination was poor, if existent. His experiment was criticized both with respect to task difficulty and inadequately controlled sound sources by Anrep (**5**), who obtained a differentiation, using the conditioned salivary response method, between 637.5 cycles and 680 cycles. Dworkin (**35**) trained two dogs to make a differentiation of about one-third of a tone, which was consistent with Anrep's earlier study, and with the previous findings of Andreyev (**4**) and of the Russian investigations summarized by Razran and Warden (**124**). In this study Dworkin compared the results from a conditioned salivary response with those from a conditioned alimentary-motor response, and found them to be substantially equivalent, although the specific responses differed somewhat in latency, duration, initial activity, and inhibition.

The upper limit of the dog's frequency spectrum was placed at 36 kilocycles to 38 kilocycles by Andreyev (**4**), using a conditioned salivary response, and somewhat higher by workers in Pavlov's laboratory (**120**, p. 127). Sutherland and Dworkin (**146**) reported that their animals responded to sound from 50 cycles to 20 kilocycles, and in one case to

30 kilocycles. They demonstrated that dogs can differentiate between air-borne sound stimuli in the range 50 cycles to 300 cycles and tactual vibrations (when the table on which the dog is placed is set into vibration) at the same frequency.

Audiograms have been determined by Lipman and Grassi (**98**) and Dworkin, Katzman, Hutchinson, and McCabe (**36**). While the results are not completely consistent at all frequencies (the dog's thresholds reported by Lipman and Grassi are appreciably lower, from 200 cycles to 4,000 cycles, than those reported by Dworkin and his associates), the studies agree that human and dog thresholds were substantially the same below 250 cycles, and that the dog's maximum sensitivity was in the range of 4,000 cycles and above. Dworkin reported a superiority of 25 decibels at 14 to 16 kilocycles. Brogden and Culler (**17**) noted that the dog's auditory acuity was increased following roentgen-radiation of the hypophysis; the effect was due to induced hypoglycemia. The increase in acuity was most striking at the higher frequencies (8,000 cycles).

Guinea pigs. Herrington and Gundlach (**72**) reported definite evidence of discrimination in the guinea pig between tones of 500 and 600 cycles, and evidence of possible discrimination between 591 and 614 cycles. Their results were not conclusive, however, since they were not careful to use pure tones. Upton (**150**) obtained a discrimination between 600 cycles and 1,000 cycles, using a conditioned respiratory response. The discrimination required between 500 and 600 trials to establish. Horton (**77**) determined the animal's (seven animals) audiogram from 64 cycles to 8,192 cycles, using the thresholds of twenty human beings "tested under the same conditions" as a standard. He also employed a conditioned respiratory response. He found that the animal's threshold was higher than the human threshold except for a few cases at middle frequencies. Horton concluded that "the capacity of these animals in the perception of sound is similar to that of man." Analysis of his results in terms of the average of the seven-animal thresholds at each frequency suggests, however, that the guinea pig's discrimination capacity may be substantially poorer than man's at all but the middle frequencies.

Rats. The early negative results of Yerkes (**180**) and Hunter (**78, 79, 80, 81**) relative to the inability of the rat to differentiate pure tone are summarized in detail in Munn (**105**, pp. 140–145) as well as the first positive results of Muenzinger and Gentry (**102**) and Thuma (**148**). Both of these workers, employing electrically driven oscillators and simple mazes, demonstrated positive discrimination between a pure tone and silence. The history of the issue is brought up to date in Munn's recent volume (**107**, pp. 158–165). Since Wada (**152**), after careful study of the rat's inner ear, concluded that the animal possessed no peculiarity of structure that should result in lack of tonal sensitivity, the later results seem to be more consistent with the facts. The difficulty Hunter encountered may be attributed to the difficulty of the task and inadequacies in the sound source used. Furthermore, the recent evidence suggests that the rat can successfully discriminate only relatively intense tones over 1,000 cycles; the apparatus used by Hunter was limited with respect to intensity of sound generated, or frequency, or both. The same difficulties probably account for Roth's (**133**) negative results with the hamster.

Several workers have determined the audiogram of the rat. Henry (**70**) used a discrimination apparatus to obtain thresholds for four frequencies between 1,000 cycles and 8,000 cycles. He noted that 1,000 cycles approached the lower limit of the rat's response to a pure tone, and that the animal's threshold is lower than the human threshold only at 8,000 cycles. At 1,000 cycles it required considerable forcing to get the animal to make a choice. Using conditioning methods, audiograms extending to higher frequencies (about 10,000 cycles) were determined by Blackwell and Schlosberg (**12**), Cowles and Pennington (**24**), Eccher (**38**), and Gould and Morgan (**51**). These findings are generally quite consistent with one another, and the trend is consistent with that noted by Henry. However, all the conditioning experiments yielded appreciably lower threshold values. The threshold is high at 1,000 cycles or less, and gradually declines until it is lower than the human threshold at 10,000 cycles.

It is likely that the upper limit of the rat's frequency response has not yet been recorded. Davis (**27a**), in 1927, reported that the harvest mouse responded to a 35,000-cycle tone generated by a Galton whistle. Gould and Morgan (**50**) report that the rat is sensitive to 40,000 cycles, and probably beyond.

Britt (**15, 16**) obtained differential conditioned responses to frequencies of 480 and 2,048 cycles, and to frequencies of 960 and 2,048 cycles, while Blackwell and Schlosberg (**12**) conditioned one rat to discriminate between 9,000 and 10,000 cycles, and another to discriminate between 8,500 cycles and 10,000 cycles. The evidence, therefore, indicates that rats differentiate between tones and silence over a wide frequency range, extending from about 1 kilocycle to far beyond the human frequency response limit, and that they differentiate tones varying in frequency.

Action Potentials and Auditory Capacity

The cochlear response has been measured for a few animals as a possible index to hearing. For the guinea pig and cat, Wever and Bray (**175, 176**) and Wever and Lawrence (**177**) found that the cochlear response bears a precise linear relationship to stimulus intensity over a wide range. The curves vary in form, in departure from linearity, and in the maximum cochlear response attained. Below 1,000 cycles, larger cochlear responses are attained with increase in stimulus intensity, than with frequencies above 1,000 cycles, where a plateau in cochlear response begins at lower stimulus intensity values (**177**). Wever and Bray (**175, 176**) and Cowell and Black (**23**) deny, however, that cochlear response levels are to be identified with loudness levels. Galambos (**46**) obtained cochlear potentials from the rat up to 98 kilocycles (the limit of the equipment) and "not above 40" kilocycles from the guinea pig. The precise relationship of cochlear response levels to acuity threshold levels is not yet clear.

Neff (**109a, 110**) conditioned cats to pure tones, obtained audiograms, and then partially sectioned the auditory nerve. Audiograms on the operated animals revealed that while very small lesions resulted in no loss, larger lesions resulted in high tone loss or in loss at all frequencies. On the other hand, Wever and Neff (**178**) found, after partial section of the eighth nerve, marked high tone loss in only one operated animal, as measured by the cochlear potential.

The foregoing summary of the auditory capacities of infrahuman mammals suggests that a certain body of significant data has been accumulated. The range of capacity in most species is still to be explored, particularly with larger populations that will permit the determination of reliable averages. Problems arising out of the conditioned response method of determining audiograms must be answered. No reliable data seemed to be available for such common animals as the horse, sheep, cow, pig. Questions such as the effect of domestication on animals have not been resolved, and only a beginning has been made in the correlation of sensory capacity with such contingent behavior as perception and learning. It is clearly possible, however, to design experiments which will take into account the essential auditory peculiarities of the species for which data are given above.

IV. GUSTATION

Parker, in his model monograph in 1922, pointed out that the comparative physiology of taste in vertebrates was an almost untouched field (**114**, p. 166). Although many experiments since then have dealt with various aspects of gustatory sensitivity, as yet few generalizations of significance have been obtained from them (**101**, p. 156). Scarcely a score of papers exist that present relatively reliable evidence about the nature and limits of infrahuman mammalian taste sensitivities. These few provide fragmentary data concerning taste thresholds for a limited number of substances for perhaps two hundred animals primarily from five species: the chimpanzee (**117**), monkey (**13, 14, 117, 119**), dog (**138**), rabbit (**43**), and rat (**7, 117, 122, 126, 127, 128, 129, 130, 133, 165, 170**). Furthermore, differences in treatment of data and manner of presenting results make interspecies and interstudy comparisons difficult. Brief reviews touching in part on infrahuman gustatory capacities have been written by Geldard (**47**), Patton (**116**), and Patton and Ruch (**118**). The fullest current account is perhaps that of Moncrieff (**100**).

Methods

The more precise among the experimental procedures employed to explore the limits of auditory and visual capacities have not yet been adapted to the determination of gustatory *sensitivity thresholds*, or the minimum stimulus values to which the animal can respond or which it is able to differentiate. Patton (**117**) has pointed out that no simple way of making reward and punishment depend upon correct discrimination of two taste thresholds has yet been invented.

The typical technique employed has involved determination of the concentration of test solution most frequently preferred by the experimental animals. This procedure, the *preference method* (**7, 13, 14, 43, 117, 119, 122, 126, 127, 128, 129, 130, 170, 187, 188**), has yielded *acceptance thresholds* (the maximum concentration of a noxious substance accepted as frequently as a neutral one, usually water, or the minimum concentration of an acceptable substance reliably preferred to a neutral one, such as water) and *rejection thresholds* (the minimum concentration of a substance consistently rejected in favor of a neutral substance such as water).

The preference threshold experiments have almost universally involved substances in solution, for they permit ease of measurement and control of texture.

In the typical experiment, a rack of funnels, water dishes, or other drinking vessels is placed in the cage. Two funnels or dishes have been used with rats, mice, rabbits; apparatus for the monkey has included as many as six, and for the chimpanzee, eight (**117**). The animal may be allowed to become accustomed to the experimental situation over a period ranging up to two weeks (**125, 126**), while records of normal water intake and solution evaporation are maintained. Most of the studies have used distilled water for the control solution, but Weddell (**170**) used tap water on the hypothesis that distilled water would "create a preference for quinine" (**170**). It may be noted that his threshold for the rat is almost twice the rejection threshold determined by Patton and Ruch (**119**).

After habituation to the apparatus, the experimental regimen is begun. The animal, which is kept on a standard maintenance schedule, is usually starved from early one evening until the following morning, when two solutions, water and a controlled concentration of the test solution, are placed in the cage. The concentration of the test solution is increased (or decreased) in small steps daily, until a threshold criterion is reached and passed. Daily records are maintained of water and test solution intake. If an ascending series is used initially, a descending series may follow; or the reverse procedure may be employed.

In an earlier and less well-controlled variation of the preference method, Watson (**165**) presented the animal with small cubes of bread soaked in a test solution. This method yielded much higher threshold values than the solution technique. Roth (**133**) offered hungry mice a choice of pairs of piles of grain dipped into various solutions. In the comparisons, the members of the pair were usually dipped into different solutions (e.g., saccharine solution and acetic acid). His method therefore gave values relative to the substances, both of which might have been equally noxious.

Young (**186**) has pointed out that preference experiments tell us what concentrations are consistently selected, but that this selection may depend upon a position habit, a food habit, or the palatability of the food. These environmental or organismic factors may be only indirectly related to the animals' gustatory discrimination capacity. Bare (**7**) noted that position habits affected the choices made by the experimental animals, and consequently the threshold determinations. Further, more fluid was ingested during the last six hours of darkness than during other periods. Beebe-Center (**9**) found that greater amounts of saccharine solution than water are ingested in the preference situation.

That preference experiments do not yield minimum sensation thresholds has been clearly indicated by studies by Bare (**7**) and Richter (**125, 126**) of the effect of adrenalectomy in the rat on the salt preference threshold. Whereas the preoperative thresholds were about 0.06 per cent, the postoperative thresholds were between 0.016 per cent (**7**) and 0.003 per cent (**126**). Richter (**126**) concluded that changes in the oral cavity resulting from the effects of adrenalectomy led to enhanced discrimination. However, Bare (**7**) found that the receptors were *not* affected: the concentrations of sodium chloride required to evoke action potentials in the chorda tympani were not significantly different in normal and adrenalectomized animals. Further, Pfaffman and Bare (**122**) found that systematic surgery on the lingual and glossopharynnerves, leading to complete degeneration of the lingual taste buds 16), resulted in slightly elevated thresholds, but in an unchanged maximal concentration preference. They concluded that some sensory mechanism other than lingual taste is essential in initiating preference behavior. Finally, Young (187) obtained preference thresholds in normal rats of between 0.05 per cent and 0.01 per cent, depending on conditions, and further found that adrenalectomized rats do not ingest a significantly greater total amount of salt than do normal rats (188).

The matter has been summed up by Patton (116, 118), who pointed out that preference thresholds differ from absolute taste thresholds. He suggested that the fact that normal rats (as compared with adrenalectomized animals) do not prefer weak salt solutions to water may only mean that motivation to discriminate is insufficient. The salt deprivation brought about by adrenalectomy may then provide the motivation for finer discrimination.

Richter and Campbell (128) noted that the preference threshold of rats for sugar (0.5 per cent) was near the human *sugar recognition threshold* (0.41 per cent), but higher than the human *difference threshold* (0.17 per cent). Richter also found that the human salt recognition threshold was 0.087 per cent (cf. rat preference threshold of 0.055 per cent), while the human difference threshold was 0.016 per cent. In view of all the foregoing, however, it would be hazardous to conclude that animal preference thresholds are in fact functionally equivalent to human taste thresholds. The evidence suggests that, given adequate motivation (as in a reward and punishment situation), rat difference thresholds may be sensibly lower than any yet observed.

More direct methods of determining gustatory thresholds have also been used, but a translation into the terms of the preference experiments is hazardous without clearer evidence than is presently available of the correlation between neural and physiological functions on the one hand and taste sensitivities on the other. Both Pfaffman (121) and Zotterman (189) explored the gustatory sensitivity of the cat by recording action potentials from the chorda tympani and glossopharyngeal nerve following application of stimulus solutions to the tongue, and concluded that cats do not seem to discriminate sugar solutions. Pfaffman and Bare (122) found, however, that while salt *preference thresholds* were elevated somewhat after severing the glossopharyngeal and lingual nerves of the rat, the maximal preference concentration in operated animals was the same as it was in normals, even though there was complete degeneration of the lingual taste buds (116). Their conclusion that preferential taste behavior is initiated by other than the lingual

TABLE III

MAMMALIAN GUSTATORY THRESHOLDS*

(In grams per 100 cc., or in per cent solution)

Animal	*N*	*Substance*	*Acceptance Threshold*	*Rejection Threshold*
Chimpanzee (117)	1	Quinine	0.0003	0.0062
Monkey (13)	3	Quinine	0.0031	0.025
Monkey (119)	2	Quinine	0.0004	0.0015
Monkey (117)	15	Quinine	0.0031	0.025
Rat (117)	9	Quinine	0.00005	0.0003
Rat (170)	14	Quinine		0.0007%
Rabbit (43)	10	Sucrose	1.061	
Rat (128)	13	Sucrose	0.524%	
Rat (129)	7–10	Sucrose	0.57%	
Rat (187)	10	Sucrose	0.5%	80%
Rat (127)	17	Alcohol	1.8–4.8%	6%
Rat (129)	7–10	Maltose	0.06%	
Rat (129)	7–10	Galactose	1.6%	
Rat (129)	7–10	Glucose	0.2%	
Rat (129)	7–10	Lactose		2.2%
Rat (126)	20	NaCl	0.055%	
Rat (7)	10	NaCl	0.06%	
Rat (187)	10	NaCl	0.01–0.05%	
Rabbit (43)	10	NaCl		8.18
Rabbit (43)	10	LiCl		7.63
Rabbit (43)	10	$CaCl_2$		7.70
Rabbit (43)	10	KCl		8.20
Rabbit (43)	10	NH_4Cl		5.35
Rabbit (43)	10	Sodium acetate		10.67
Rabbit (43)	10	Sodium sulfate		9.23
Rabbit (43)	10	Sodium bromide		9.26
Rabbit (43)	10	Sodium nitrate		6.89
Rabbit (43)	10	Sodium iodide		11.84
Dog (138)	2	Lactic acid		0.83%
Dog (138)	2	Hydrochloric acid		0.31%
Rabbit (43)	10.	Hydrochloric acid		0.437
Rabbit (43)	10	Acetic acid		1.02
Dog (138)	2	Acetic acid		0.156%
Rabbit (43)	10	Sulfuric acid		0.687
Rabbit (43)	10	Nitric acid		0.054
Rat (130)	47	Phenylthiocarbamide		0.0003%

* The original data in **13**, **117**, and **119** were in terms of gr/100 cc; in **43** they were in terms of normal solutions. Conversion of per cent solutions to gr/100 cc (or to normal solutions) is not readily achieved, since it would require knowing the volume occupied by the solute. Under, say, 0.5 per cent, the per cent figure will be only very slightly greater in absolute value than the gr/100 cc. At higher per cent values, the discrepancy will be relatively great depending on the density of the solute. The normal solutions were therefore converted to gr/100 cc, and the per cent values were included as given in the original reports.

apparatus has been cited above. Schellworth (**138**) made a study of the dog's sensitivity by noting, in animals with duodenal fistulae, the smallest concentrations of acids mixed with meat broth and bread that resulted in the pylorus reflex.

Treatment of Data

The preference technique has been characteristically employed with the psychophysical method of limits. However, calculation of the acceptance-rejection ratio and setting of criterion points has differed from experiment to experiment. Patton, Blum, Ruch, and Walker have determined the relative preference at each concentration of the test solution (t.s.) by the ratio

$$\frac{\text{total t.s.}}{\text{total t.s.} + \text{total water}} \times 100,$$

and they have apparently set the acceptance threshold at a ratio of 50 per cent, and the rejection threshold at or near zero per cent (**13**, **117**, **119**). Others, however, have used the ratio

$$\frac{\text{total t.s.}}{\text{total water}}$$

(**125**, **127**, **128**, **129**, **170**). With this technique, the threshold has been placed at the concentration at which the ratio is 1.0. Weddell (**170**) found that the ratio varied from 1.2 to 0.4 when both dishes contained water, and he therefore placed the threshold at the weakest concentration above a ratio of 1.9. With this criterion, he recorded, for 14 rats, an average threshold for quinine of 0.000718 per cent (while Patton and Ruch [**117**] set the threshold at 0.0003 gm/100cc, approximately half as strong a solution), and a range from 0.001125 per cent to 0.000375 per cent. In view of this spread of individual variability (which is not unusual—e.g., Patton and Clisby [**130**] found the preference threshold for phenylthiocarbamide for rats to vary from 0.02 per cent to 0.000005 per cent), and in view of the differences of calculation methods and criteria, current threshold data must be viewed with caution.

Gustatory Thresholds

Table III presents the thresholds for a variety of substances for the chimpanzee, monkey, rat, rabbit, and dog. The various investigators used different bases for expressing solution concentrations; the conversions effected here are only approximately correct.

It is evident from these data that meaningful generalizations concerning the gustatory sensitivities of infrahuman mammals are virtually impossible. Too few animals and too few substances have been tested. Acceptance and rejection thresholds (as compared with sensory thresholds) have been demonstrated to be peculiarly susceptible to variation with intercurrent metabolic needs and maintenance schedules of different characteristics, and can but be ambiguous. Conclusions with respect to the relative sensitivities of man and animals, where human thresholds are almost invariably sensation thresholds (cf. **43**, **117**, **128**), are open to question. It seems quite likely that, in any given animal, the sensation threshold will be lower by some amount than either the rejection or acceptance threshold. Further, the range of intraspecies differences is such that the means from small samples are likely to be quite unstable.

V. SUMMARY

The present article has presented a summary of the existing knowledge with respect to certain of the visual, auditory, and gustatory capacities of infrahuman mammals. While no pretense to completeness is made, it is believed that the larger part of the reliable evidence has been included. This survey has brought out three important points:

1. Only a few of the numerous infrahuman mammals have been studied at all, and of these only one or two have been studied with any degree of intensity.
2. The small sizes of samples employed, considered in the light of the usually large intraspecies variation, call into question the reliability of many of the findings.
3. Lack of comparability of methods used to explore the same capacity in the same or in different species makes it difficult to compare the results secured in these studies.

This paper has been concerned primarily with visual acuity, color vision, auditory frequency thresholds, and gustatory discrimination. It seems worth while briefly to compare the status of work in these fields.

1. It is the writer's judgment that the most reliable measurements are available for the largest range of animals with respect to audition. Consistent audiograms, based on essentially the same technique, have been obtained for the rat, dog, cat, and infrahuman primate. Both experimental method and instrumentation have developed to a point where some confidence can be placed in reported findings.
2. With respect to color vision, the findings reported are largely qualitative. In most instances it is safe to conclude only that the animal can discriminate some or all hues. The significant recent work has concerned only the rat and primates. It seems likely that the scattered early reports on other animals, such as the raccoon, porcupine, cat, guinea pig, squirrel, should be largely discounted.

Both the experimental methods and color sources used were not adequate by current standards. The history of color vision research with the rat suggests that negative findings particularly should be viewed with extreme caution. While color vision research accounts for a larger number of papers than any other area, only a small fraction of these are in any sense definitive.

3. As for visual acuity, not more than a score of individual animals seem to have been studied, and visual acuity in these was studied with a half-dozen different techniques, under as many different conditions. Both method and equipment are available, however, to undertake intensive comparative studies.

4. The area of gustation is difficult to evaluate. The preference technique, which has been employed almost exclusively, yields a measurement that cannot be legitimately construed as an index of minimum capacity to discriminate. Infrahuman mammalian audiograms, for example, are much more comparable to human audiograms than infrahuman preference thresholds are to human taste thresholds, even if it may be necessary to discount the effect of intensive practice with respect to the obtained audiograms. There are more or less *consistent* measurements available for one animal, the rat, for at least sodium chloride and sucrose. There are a few possibly significant measurements for the rabbit, and for primate sensitivity to quinine. More than in any other area, there is a need for a technique for gustatory research that will make possible determination of the *limits* of the animal's capacity to discriminate.

5. Finally, in no area considered has the problem of the relationship between the receptor mechanism and response capacity been more than broached. The locus of auditory threshold discrimination and of taste discrimination are both in doubt. The conflicting evidence with respect to the anatomy of the eye and color vision still remains to be resolved.

On the positive side, however, it may be noted that both stimulus-generating equipment and research techniques have been greatly improved and refined in the last ten years. The facilities are available for a full-scale attack on the problems adumbrated above.

The writer believes that such an attack is necessary. The essential reason for the study of comparative psychology is methodological: to study as completely as possible all the factors which underlie behavior, and to test and determine under controlled experimental conditions all the variables intervening between the presentation of stimuli and the measurement of responses. For many problems, the required degree of control can be achieved *only* with animals. However, before studying the "higher-order" problems of learning, perception, motivation, it is necessary to know, or make assumptions about, the contingent sensory response capacities of the subjects used.

In an experiment on human motivation, if we knew our subjects were blind, it is not likely that we would depend upon visual means of communication. It is not improbable, however, that we have made such mistakes in working with animals. It was the intent of this paper to point out the extent of our ignorance of the basic sensory capacity characteristics of animals, and to add a mite of motivation to attack the problems outlined.

BIBLIOGRAPHY

1. Ades, H. W. Effect of lesions in the medial geniculate bodies upon hearing in the cat. *Amer. J. Physiol.*, 1939, **125**, 15–23.
2. Ades, H. W., Mettler, F. A., & Culler, E. A. Distribution of acoustic pathways in the medial geniculate bodies. *Amer. J. Physiol.*, 1937, **119**, 257–258. (Abstract.)
3. Allen, Jessie. Associative processes in the guinea pig. *J. comp. Neurol. & Psychol.*, 1904, **14**, 293.
4. Andreyev, L. A. Extreme limits of pitch discrimination with higher tones. *J. comp. Psychol.*, 1934, **18**, 315–332.
5. Anrep, G. V. Pitch discrimination in the dog. *J. Physiol.*, 1920, **53**, 367–385.
6. Bárány, E. The special senses. Part I. Hearing. *Ann. Rev. Physiol.*, 1941, **3**, 449–460.
7. Bare, J. K. The specific hunger for sodium chloride in normal and adrenalectomized white rats. *J. comp. physiol. Psychol.*, 1949, **42**, 242–253.
8. Bartley, S. H. *Vision*. New York: Van Nostrand, 1941.

8a. Bartley, S. H. Sense organs. Part I. Vision. *Ann. Rev. Physiol.*, 1943, **5**, 345–356.

9. Beebe-Center, J. G., Black, P., Hoffman, A. C., & Wade, Marjorie. Relative per diem consumption as a measure of preference in the rat. *J. comp. physiol. Psychol.*, 1948, **41**, 239–251.
10. Bierens de Haan, J. A. Experiments on vision in monkeys. I. The colour-sense of the pig-tailed macaque (*Nemestrinus nemistrinus L.*) *J. comp. Psychol.*, 1925, **5**, 417–455.
11. Bierens de Haan, J. A., & Frima, Margareta J. Versuche über den Farbensinn der Lemuren. *Z. vergl. Physiol.*, 1930, **12**, 603–631.
12. Blackwell, H. R., & Schlosberg, H. Octave generalization, pitch discrimination, and loudness thresholds in the white rat. *J. exp. Psychol.*, 1943, **33**, 407–419.
13. Blum, M., Walker, H. E., & Ruch, T. C. Localization of taste in the thalamus of *Macaca mulatta*. *Yale J. Biol. Med.*, 1943, **16** 175–191.
14. Börnstein, W. S. The cortical taste area in monkeys and a semi-quantitative method of testing taste in monkeys. *Amer. J. Physiol.*, 1940, **129**, 314. (Abstract.)
15. Britt, S. Tonal sensitivity in the white rat. *Psychol. Bull.*, 1934, **31**, 619. (Abstract.)
16. Britt, S. H. Tonal sensitivity in the white rat. *J. comp. Psychol.*, 1935, **19**, 243–264.
17. Brogden, W. J., & Culler, E. A. Increased acoustic sensitivity in dogs following roentgen-radiation of the hypophysis. *Amer. J. Physiol.*, 1937, **119**, 13–23.
18. Brown, R. H. The brightness visibility curve of the rabbit. *J. gen. Psychol.*, 1937, **17**, 323–338.
19. Cole, L. W. Concerning the intelligence of racoons. *J. comp. Neurol. & Psychol.*, 1907, **17**, 211–261.
20. Cole, L. W., & Long, F. M. Visual discrimination in racoons. *J. comp. Neurol. & Psychol.*, 1909, **19**, 83–107.
21. Coleman, T. B., & Hamilton, W. F. Color blindness in the rat. *J. comp. Psychol.*, 1933, **15**, 177–181.
22. Colvin, S. S., & Burford, C. C. Color perception of three dogs, a cat and a squirrel. *Psychol. Monogr.*, 1909, **11**, No. 44.
23. Covell, W. P., & Black, L. J. The cochlear response as an index to hearing. *Amer. J. Physiol.*, 1936, **116**, 524–530.
24. Cowles, J. T., & Pennington, L. A. An improved technique for determining auditory acuity of the rat. *J. Psychol.*, 1943, **15**, 41–47.
25. Crawford, M. P., & Brundage, E. G. Recent methods for generating

sound stimuli for use in testing the auditory capacity in animals. *Quart. Rev. Biol.*, 1932, **7**, 444–457.

26. Culler, E. A., Finch, G., Girden, E., & Brogden, W. J. Measurements of acuity by the conditioned-response technique. *J. gen. Psychol.*, 1935, **12**, 223–227.

27. Culler, E. A., Willmann, J., & Mettler, F. A. Mapping the cochlea. *Amer. J. Physiol.*, 1937, **119**, 292. (Abstract.)

27a. Davis, A. C. Sound perception in a harvest mouse. *J. Mammal.*, 1927, **8**, 161–162.

28. Davis, H. B. The racoon. *Amer. J. Psychol.*, 1907, **18**, 447–489.

29. Davis, H. Sense organs. Part II. Special senses other than vision. *Ann. Rev. Physiol.*, 1943, **5**, 357–372.

30. Detwiler, S. R. *Vertebrate photoreceptors.* New York: Macmillan, 1943.

31. DeVoss, J. C., & Ganson, R. Color blindness of cats. *J. Anim. Behav.*, 1915, **5**, 115–139.

32. Drury, D. W. Tests of hearing of five hundred average ears by the audiometer No. 2-A. *Arch. Otolaryng.*, 1925, **1**, 524–532.

33. Dworkin, S. Hearing tests in normal and operated dogs and cats. *Trans. Amer. otol. Soc.*, 1934, **24**, 143–151.

34. Dworkin, S. Pitch and intensity discrimination by cats. *Amer. J. Physiol.*, 1935, **112**, 1–4.

35. Dworkin, S. Alimentary motor conditioning and pitch discrimination in dogs. *Amer. J. Physiol.*, 1935, **112**, 323–328.

36. Dworkin, S., Katzman, J., Hutchinson, G. A., & McCabe, J. R. Hearing acuity of animals as measured by conditioning methods. *J. exp. Psychol.*, 1940, **26**, 281–298.

37. Dworkin, S., Sutherland, G. F., & Seymour, S. L. A conditioned reflex method of testing hearing in cats. *Quart. J. exp. Physiol.*, 1934, **24**, 23–30.

38. Eccher, W. Comparative thresholds of pitch and intensity in rat, dog and man. Paper delivered at Eastern Psychological Association Meeting, April, 1942. (Cited in H. R. Blackwell and H. Schlosberg, Octave generalization, pitch discrimination and loudness thresholds in the white rat. *J. exp. Psychol.*, 1943, **33**, 407–419.)

39. Elder, J. H. Auditory acuity of the chimpanzee. *J. comp. Psychol.*, 1934, **17**, 157–183.

40. Elder, J. H. The upper limit of hearing in chimpanzee. *Amer. J. Physiol.*, 1935, **112**, 109–115.

41. Fields, P. E. Form discrimination in the white rat. *J. comp. Psychol.*, 1928, **8**, 143–158.

42. Fields, P. E. Discrimination. In F. A. Moss (Ed.), *Comparative psychology.* New York: Prentice-Hall, 1946. Ch. 6, pp. 130–152.

43. Frings, H. A. A contribution to the comparative physiology of contact chemoreception. *J. comp. physiol. Psychol.*, 1948, **41**, 25–34.

44. Fulton, J. F., Hoff, P. M., & Perkins, H. T. *A bibliography of visual literature, 1939–1944.* Springfield, Ill.: Charles C Thomas, 1945.

45. Furness, W. H. Observations on mentality of chimpanzees and orang-utans. *Proc. Amer. phil. Soc.*, 1916, **55**, 281–290.

46. Galambos, R. Cochlear potentials from the bat. *Science*, 1941, **93**, 215.

47. Geldard, F. A. Somesthesis and the chemical senses. *Ann. Rev. Psychol.*, 1950, **1**, 71–86.

48. Gentry, E. Methods of discrimination training in white rats. *J. comp. Psychol.*, 1934, **18**, 227–258.

49. Ghiselli, E. E., & Henry, F. M. The role of sensory acuity in learning and sensory discriminations. *J.*

genet. Psychol., 1936, **49**, 498–499.

50. Gould, J., & Morgan, C. T. Hearing in the rat at high frequencies. *Science*, 1941, **94**, 168.
51. Gould, J., & Morgan, C. T. Auditory sensitivity in the rat. *J. comp. Psychol.*, 1942, **34**, 321–329.
52. Graham, C. H., Kemp, E. H., & Riggs, L. A. An anlaysis of the electrical potential responses of a color-discriminating eye to lights of different wave-lengths. *J. gen. Psychol.*, 1935, **13**, 275–296.
53. Graham, C. H., & Riggs, L. A. The visibility curve of the white rat as determined by the electrical retinal response to lights of different wave-lengths. *J. gen. Psychol.*, 1935, **12**, 279–293.
53a. Granit, R. The special senses. Part II. Visual receptors. *Ann. Rev. Physiol.*, 1941, **3**, 461–474.
54. Granit, R. The spectral properties of the visual receptors of the cat. *Acta Phys. Sci.*, 1943, **5**, 219–229.
55. Gregg, F. M., Jamison, Eva, Wilkie, R., & Radinsky, T. Are dogs, cats and raccoons color-blind? *J. comp. Psychol.*, 1929, **9**, 379–395.
56. Gregg, F. M., & McPheeters, C. A. Behavior of raccoons to temporal series of stimuli. *J. Anim. Behav.*, 1913, **3**, 241–259.
57. Grether, W. F. A new spectral color apparatus for studies of primate color vision. *J. gen. Psychol.*, 1935, **12**, 450–456.
58. Grether, W. F. Color vision and color blindness in monkeys. *Comp. Psychol. Monogr.*, 1939, **15**, No. 76.
59. Grether, W. F. Chimpanzee color vision. I. Hue discrimination at three spectral points. *J. comp. Psychol.*, 1940, **29**, 167–178.
60. Grether, W. F. Chimpanzee color vision. II. Color mixture proportions. *J. comp. Psychol.*, 1940, **29**, 179–186.
61. Grether, W. F. Chimpanzee color vision. III. Spectral limits. *J. comp. Psychol.*, 1940, **29**, 187–192.
62. Grether, W. F. A comparison of human and chimpanzee spectral hue discrimination curves. *J. exp. Psychol.*, 1940, **26**, 394–403.
63. Grether, W. F. Comparative visual acuity thresholds in terms of retinal image widths. *J. comp. Psychol.*, 1941, **31**, 23–33.
64. Grether, W. F. Spectral saturation curves for chimpanzee and man. *J. exp. Psychol.*, 1941, **28**, 419–427.
65. Harlow, H. F. Studying animal behavior. In T. G. Andrews (Ed.), *Methods of psychology*. New York: Wiley, 1948. Ch. 12, pp. 319–347.
66. Harris, J. D. The auditory acuity of pre-adolescent monkeys. *J. comp. Psychol.*, 1943, **35**, 255–265.
67. Hartline, H. K. Sense organs. *Ann. Rev. Physiol.*, 1942, **4**, 445–464.
68. Henry, F. M. Comparative thresholds of audition in the white rat. *Psychol. Bull.*, 1935, **32**, 536. (Abstract.)
69. Henry, F. M. Audition in the white rat. II. The learning of a pure tone discrimination. *J. comp. Psychol.*, 1936, **22**, 105–121.
70. Henry, F. M. Audition in the white rat. III. Absolute and relative intensity thresholds. *J. comp. Psychol.*, 1938, **26**, 45–62.
71. Henry, F. M., & Brown, C. W. Audition in the white rat. I. A pure tone discrimination apparatus of high reliability. *J. comp. Psychol.*, 1936, **21**, 1–8.
72. Herington, G. B., & Gundlach, R. H. How well can guinea pigs and cats hear tones? *J. comp. Psychol.* 1933, **16**, 287–303.
73. Honzig, C. H. The sensory basis of maze learning in rats. *Comp. Psychol. Monogr.*, 1936, **13**, No. 64.
74. Hopkins, A. E. Experiments on color vision in mice in relation to the

duplicity theory. *Z. vergl. Physiol.*, 1927, **6**, 299–344.

75. Hopkins, A. E. Vision and retinal structure in mice. *Proc. nat. Acad. Sci., Wash.*, 1927, **13**, 488–491.

76. Hopkins, A. E. Vision in mice with "rodless" retinae. *Z. vergl. Physiol.*, 1927, **6**, 345–360.

77. Horton, G. P. A quantitative study of hearing in the guinea pig (*Cavia cobaya*). *J. comp. Psychol.*, 1933, **15**, 59–73.

78. Hunter, W. S. The auditory sensitivity of the white rat. *J. Anim. Behav.*, 1914, **4**, 215–222.

79. Hunter, W. S. The auditory sensitivity of the white rat. *J. Anim. Behav.*, 1915, **5**, 312–329.

80. Hunter, W. S. Some notes on the auditory sensitivity of the white rat. *Psychobiology*, 1918, **1**, 339–351.

81. Hunter, W. S. Further data on the auditory sensitivity of the white rat. *Pedagog. Sem. & J. genet. Psychol.*, 1927, **34**, 177–187.

82. Hunter, W. A., & Pennington, L. A. A new apparatus and method for training the rat in auditory discrimination problems. *Science*, 1939, **89**, 87–88.

83. Hunter, W. A., & Pennington, L. A. The construction and use of a new apparatus in the training of the rat in auditory discrimination problems. *J. genet. Psychol.*, 1940, **57**, 451–460.

83a. Johnson, H. M. Audition and habit formation in the dog. *Behav. Monogr.*, 1913, **2**, 1–78.

84. Johnson, H. M. Visual pattern discrimination in vertebrates. I. Problems and methods. *J. Anim. Behav.*, 1914, **4**, 319–339.

85. Johnson, H. M. Visual pattern discrimination in vertebrates. II. Comparative visual acuity in the dog, the monkey, and the chick. *J. Anim. Behav.*, 1914, **4**, 340–361.

86. Johnson, H. M. Visual pattern discrimination in vertebrates. III. Effective differences in width of visible striae for the monkey and the chick. *J. Anim. Behav.*, 1916, **6**, 169–188.

87. Johnson, H. M. Visual pattern discrimination in vertebrates. IV. Effective differences in the direction of visible striae for the monkey and the chick. *J. Anim. Behav.*, 1916, **6**, 189–204.

88. Kinnaman, A. J. Mental life of two *Macacus rhesus* monkeys in captivity. *Amer. J. Psychol.*, 1902, **13**, 98–148, 173–218.

89. Kittredge, E. Some experiments on the brightnesss value of red for the light-adapted eye of the calf. *J. comp. Psychol.*, 1923, **3**, 141–145.

90. Klüver, H. *Behavior mechanisms in monkeys.* Chicago: Univ. Chicago Press, 1933.

91. Köhler, W. Die Farbe der Sehdinge beim Schimpansen und beim Haushuhn. *Z. Psychol.*, 1917, **77**, 248–255.

92. Köhler, W. Nachweis einfacher Strukturfunktionen beim Schimpansen und beim Haushuhn; über eine neue Methode zur Untersuchung des bunten Farbensystems. *Abh. Koenig preuss. Akad. Wiss., phys.-math. Klasse*, 1918, No. 2, 1–101.

93. Kuroda, A. Mental life of a *Macacus* monkey. *Jap. J. Psychol.*, 1926, **6**, 148–174. (*Psychol. Abstr.*, 1927, **1**, 1099.)

94. Kuroda, R. Experimental researches upon the sense of hearing in lower vertebrates, including reptiles, amphibians, and fishes. *Comp. Psychol. Monogr.*, 1926, **3**, No. 16.

95. Lashley, K. S. The mechanism of vision. I. A method for the rapid analysis of pattern-vision in the rat. *J. genet. Psychol.*, 1930, **37**, 453–460.

96. Lashley, K. S. The mechanism of vision. III. The comparative vis-

ual acuity of pigmented and albino rats. *J. genet. Psychol.*, 1930, **37**, 481–484.

97. Lashley, K. S. The mechanism of vision. V. The structure and image-forming power of the rat's eye. *J. comp. Psychol.*, 1932, **13**, 173–200.
98. Lipman, E., & Grassi, J. R. Comparative auditory sensitivity of man and dog. *Amer. J. Psychol.*, 1942, **55**, 84–89.
99. Maier, N. R. F., & Schneirla, T. C. *Principles of animal psychology.* New York: McGraw-Hill, 1935.
100. Moncrieff, R. W. *The chemical senses.* New York: Wiley, 1946.
101. Morgan, C. T. *Physiological psychology.* New York: McGraw-Hill, 1943.
102. Muenzinger, K. F., & Gentry, E. Tone discrimination in white rats. *J. comp. Psychol.*, 1931, **12**, 195–206.
103. Muenzinger, K. F., & Reynolds, H. E. Color vision in white rats. I. Sensitivity to red. *J. genet. Psychol.*, 1936, **48**, 58–71.
104. Munn, N. L. An investigation of color vision in the hooded rat. *J. genet. Psychol.*, 1932, **40**, 351–362.
105. Munn, N. L. *An introduction to animal psychology.* Boston: Houghton Mifflin, 1933.
106. Munn, N. L. Further evidence concerning color blindness in rats. *J. genet. Psychol.*, 1934, **45**, 285–302.
107. Munn, N. L. *Handbook of psychological research on the rat.* Boston: Houghton Mifflin, 1950.
108. Munn, N. L., & Collins, M. Discrimination of red by white rats. *J. genet. Psychol.*, 1936, **48**, 72–87.
109. Neet, C. C. A comparison of verbal, manual, and conditioned-response methods in the determination of auditory thresholds. *J. exp. Psychol.*, 1936, 19, 401–416.

109a. Neff, W. D. An experimental investigation of hearing following partial section of the VIIIth nerve. *Psychol. Bull.*, 1940, **37**, 527–528. (Abstract.)

110. Neff, W. D. The effects of partial section of the auditory nerve. *J. comp. physiol. Psychol.*, 1947, **40**, 203–215.
111. Newman, E. B. Hearing. *Ann. Rev. Psychol.*, 1950, **1**, 49–70.
112. Olmstead, J. M. D. The special senses. *Ann. Rev. Physiol.*, 1940, **2**, 287–308.
113. Olmstead, J. M. D. The special senses. *Ann. Rev. Physiol.*, 1945, **7**, 509–526.
114. Parker, G. H. *Smell, taste, and allied senses in the vertebrates.* Philadelphia: Lippincott, 1922.
115. Parsons, J. H. *An introduction to the study of color vision.* Cambridge, Eng.: Cambridge Univ. Press, 1924.
116. Patton, H. D. Physiology of smell and taste. *Ann. Rev. Physiol.*, 1950, **12**, 469–484.
117. Patton, H. D., & Ruch, T. C. Preference thresholds for quinine hydrochloride in chimpanzee, monkey and rat. *J. comp. Psychol.*, 1944, **37**, 35–49.
118. Patton, H. D., & Ruch, T. C. Taste. In J. E. Fulton (Ed.), *Howell's Textbook of physiology* (15th Ed.). Philadelphia: W. B. Saunders, 1946. Ch. 18.
119. Patton, H. D., Ruch, T. C., & Walker, A. E. Experimental hypogeusia from Horsley-Clarke lesions of the thalamus in *Macaca mulatta*. *J. Neurophysiol.*, 1944, **7**, 171–184.
120. Pavlov, I. P. *Lectures on conditioned reflexes.* (Trans. by W. H. Gantt.) New York: International, 1928.
121. Pfaffmann, C. Gustatory afferent impulses. *J. cell. comp. Physiol.*, 1941, **17**, 243–258.
122. Pfaffmann, C., & Bare, J. K. Salt preference following denervation of

the tongue. *Amer. Psychologist*, 1948, **3**, 254. (Abstract.)

123. Raab, D. H., & Ades, H. W. Cortical and midbrain mediation of a conditioned discrimination of acoustic intensities. *Amer. J. Psychol.*, 1946, **59**, 59–83.
124. Razran, H. S., & Warden, C. J. The sensory capacities of the dog as studied by the conditioned reflex method (Russian schools). *Psychol. Bull.*, 1929, **26**, 202–222.
125. Richter, C. P. Increased salt appetite in adrenalectomized rats. *Amer. J. Physiol.*, 1936, **115**, 155–161.
126. Richter, C. P. Salt taste thresholds of normal and adrenalectomized rats. *Endocrinology*, 1939, **24**, 367–371.
127. Richter, C. P., & Campbell, K. H. Alcohol taste thresholds and concentrations of solutions preferred by rats. *Science*, 1940, **91**, 507–509.
128. Richter, C. P., & Campbell, K. H. Sucrose taste thresholds of rats and humans. *Amer. J. Physiol.*, 1940, **128**, 291–297.
129. Richter, C. P., & Campbell, K. H. Taste thresholds and taste preferences of rats for five common sugars. *J. Nutrition*, 1940, **20**, 31–46.
130. Richter, C. P., & Clisby, K. H. Phenylthiocarbamide taste thresholds of rats and human beings. *Amer. J. Physiol.*, 1941, **134**, 157–164.
131. Rochon-Duvigneaud, A. *Les yeux et la vision des vertébrés*. Paris: Masson, 1943.
132. Rosenzweig, M. Discrimination of auditory intensities in the cat. *Amer. J. Psychol.*, 1946, **59**, 127–136.
133. Roth, H. Psychologische Untersuchungen an nichtdomestizierten Nagetieren, namentlich der Hausmaus. *Z. Psychol.*, 1925, **97**, 62–88.
134. Ruch, T. C. *Bibliographia primatologica* (Part I). Baltimore: Charles C Thomas, 1941.
135. Russell, R. W., & Walton, W. E. The effects of avitaminosis-A on visual discrimination in the rat: Gross effects on color vision. *J. genet. Psychol.*, 1942, **61**, 327–333.
136. Sackett, L. W. Canada porcupine. *Behav. Monogr.*, 1913, **2**, No. 2.
137. Samojleff, A., & Pheophilatkowa, A. Über die Farbenwahrnehmung beim Hunde. *Zbl. Physiol.*, 1908, **22**, 133–139.
138. Schellworth, W. W. Receptionsorgane ohne Empfindung und ihr Verhältnis zu den Sinnesorganen. *Z. Biol.*, 1922, **76** 121–126.
139. Shepherd, W. T. Some mental processes of the Rhesus monkey. *Psychol. Monogr.*, 1910, **12**, No. 5.
140. Smith, E. M. Color vision in dogs. *Brit. J. Psychol.*, 1912, **5**, 119–202.
141. Smith, K. U. Apparatus for the study of discrimination in higher vertebrates. *J. genet. Psychol.*, 1933, **43**, 462–466.
142. Smith, K. U. Visual discrimination in the cat. IV. The visual acuity of the cat in relation to stimulus distance. *J. genet. Psychol.*, 1936, **49**, 297–313.
143. Smith, K. U., & Warkentin, J. The central neural organization of optic functions related to minimum visual acuity. *J. genet. Psychol.*, 1939, **55**, 177–195.
144. Spence, K. W. Visual acuity and its relation to brightness in chimpanzee and man. *J. comp. Psychol.*, 1934, **18**, 333–361.
145. Stagner, R. Visually determined reactions in the vertebrates. *Psychol. Bull.*, 1931, **28**, 99–129.
146. Sutherland, G. F., & Dworkin, S. Conditioned responses to sound and vibrations. *Amer. J. Physiol.*, 1932, **101**, 97–98.
147. Thorndike, E. L. *Animal intelligence*. New York: Macmillan, 1911.

148. Thuma, B. D. The response of the white rat to tonal stimuli. *J. comp. Psychol.*, 1932, **13**, 57–86.

149. Trendelberg, W., & Schmidt, I. Untersuchungen über das Farbensystem des Affen. *Z. vergl. Physiol.*, 1930, **12**, 249–278.

150. Upton, M. The auditory sensitivity of guinea pigs. *Amer. J. Psychol.*, 1929, **41**, 412–421.

151. Vincent, S. B. The mammalian eye. *J. Anim. Behav.*, 1912, **2**, 249–255.

152. Wada, T. Anatomical and physiological studies on the growth of the inner ear of the albino rat. *Amer. anat. Mem.*, No. 10. Philadelphia: Wistar Institute, 1923.

153. Walls, G. L. The visual cells of the white rat. *J. comp. Psychol.*, 1934, **18**, 363–366.

154. Walls, G. L. *The vertebrate eye.* Bloomfield Hills, Mich.: Cranbrook Institute of Science, 1942.

155. Walton, W. E. Color vision and color preference in the albino rat. *J. comp. Psychol.*, 1933, **15**, 359–394.

156. Walton, W. E. Exploratory experiments in color discrimination. *J. genet. Psychol.*, 1936, **48**, 221–226.

157. Walton, W. E., & Bornemeier, R. W. Further evidence of color discrimination in rodents. *J. genet. Psychol.*, 1938, **52**, 165–181.

158. Walton, W. E., & Bornemeier, R. W. Color discrimination in rats. *J. comp. Psychol.*, 1939, **28**, 417–436.

159. Walzl, E. M. Hearing. *Ann. Rev. Physiol.*, 1949, **11**, 231–244.

160. Warden, C. J., Jenkins, T. N., & Warner, L. H. *Comparative psychology: Principles and methods* (Vol. I). New York: Ronald, 1935.

161. Warden, C. J., Jenkins, T. N., & Warner, L. H. *Comparative psychology: Vertebrates* (Vol. II). New York: Ronald, 1936.

162. Warkentin, J. The visual acuity of some vertebrates. *Psychol. Bull.*, 1937, **34**, 793. (Abstract.)

163. Washburn, Margaret. *The animal mind.* (3rd Ed.) New York: Macmillan, 1926.

164. Washburn, Margaret, & Abbott, E. Experiments on the brightness value of red for the light adapted eye of the rabbit. *J. Anim. Behav.*, 1912, **2**, 145–180.

165. Watson, J. B. Kinesthetic and organic sensations: Their role in the reactions of the white rat to the maze. *Psychol. Rev. Monogr.*, 1907, **8**, No. 33.

166. Watson, J. B. Some experiments bearing upon color vision in monkeys. *J. comp. Neurol. & Psychol.*, 1909, **19**, 1.

167. Watson, J. B. *Behavior: An introduction to comparative psychology.* New York: Holt, 1914.

168. Watson, J. B., & Watson, M. I. A study of the responses of rodents to monochromatic light. *J. Anim. Behav.*, 1913, **3**, 1–14.

169. Waugh, K. T. Role of vision in mental life of mouse. *J. comp. Neurol. & Psychol.*, 1910, **20**, 549–599.

170. Weddell, C. H. The taste sensitivity of the white rat. I. Sensitivity to quinine sulphate. *J. comp. Psychol.*, 1936, **21**, 233–244.

171. Weinstein, B., & Grether, W. F. A comparison of visual acuity in the rhesus monkey and man. *J. comp. Psychol.*, 1940, **30**, 187–195.

172. Wendt, G. R. Auditory acuity of monkeys. *Comp. Psychol. Monogr.*, 1934, **10**, No. 3.

173. Wever, E. G. The upper limit of hearing in the cat. *J. comp. Psychol.*, 1930, **10**, 221–233.

174. Wever, E. G. Impulses from the acoustic nerve of the guinea pig, rabbit, and rat. *Amer. J. Psychol.*, 1931, **43**, 457–462.

175. Wever, E. G., & Bray, C. W. The

nature of acoustic response: The relation between sound intensity and the magnitude of responses in the cochlea. *J. exp. Psychol.*, 1936, **19**, 129–143.

176. Wever, E. G., & Bray, C. W. The nature of acoustic response: The relation between stimulus intensity and the magnitude of cochlear responses in the cat. *J. exp. Psychol.*, 1938, **22**, 1–16.

177. Wever, E. G., & Lawrence, M. The patterns of response in the cochlea. *J. acoust. Soc. Amer.*, 1949, **21**, 127–134.

178. Wever, E. G., & Neff, W. D. A further study of the effects of partial section of the auditory nerve. *J. comp. physiol. Psychol.*, 1947, **40**, 217–226.

179. Weymouth, F. W. Vision. *Ann. Rev. Physiol.*, 1944, **6**, 391–426.

180. Yerkes, R. M. *The dancing mouse.* New York: Macmillan, 1907.

181. Yerkes, R. M. *Chimpanzees.* New Haven: Yale Univ. Press, 1943.

182. Yerkes, R. M., & Morgulis, S. Method of Pavlov in animal psychology. *Psychol. Bull.*, 1909, **6**, 257–273.

183. Yerkes, R. M., & Petrunkevitch, A. Studies of chimpanzee vision by Ladygin-Kohts. *J. comp. Psychol.*, 1925, **5**, 99–108.

184. Yerkes, R. M., & Watson, J. B. Methods of studying vision in animals. *Behav. Monogr.*, 1911, **1**, No. 2.

185. Yerkes, R. M., & Yerkes, Ada W. *The great apes.* New Haven: Yale Univ. Press, 1929.

186. Young, P. T. Studies of food preference, appetite, and dietary habit. VI. Habit, palatability and diet as factors regulating the selection of food by the rat. *J. comp. Psychol.*, 1946, **39**, 139–176.

187. Young, P. T. Studies of food preference, appetite, and dietary habit. IX. Palatability versus appetite as determinants of the critical concentrations of sucrose and sodium chloride. *Comp. Psychol. Monogr.*, 1949, **19**, No. 15.

188. Young, P. T., & Chaplin, J. P. Studies of food preference, appetite, and dietary habit. X. Prefeences of adrenalectomized rats for salt solutions of different concentrations. *Comp. Psychol. Monogr.*, 1949, **19**, No. 15.

189. Zotterman, Y. Action potentials in the glossopharyngeal nerve and in the chorda tympani. *Skand. Arch. Physiol.*, 1935, **72**, 73–77.

190. *A bibliography in audition.* Psycho-Acoustic Laboratory, Harvard University. Cambridge, Mass.: Harvard Univ. Press, 1950.

Received August 19, 1950.

EXPERIMENTAL NAÏVETÉ AND EXPERIENTIAL NAÏVETÉ[1]

RICHARD CHRISTIE

Research Center for Human Relations, New York University

It may seem to be a truism to assert that early experience in the life of any complex organism leaves an experiential residue which affects later behavior. Students of personality have amassed a considerable amount of evidence relating to the effects of early experience upon adult behavior in homo sapiens (11). In view of this impressive array of evidence indicating the importance of early experience upon subsequent behavior, few animal experimenters would deny that pre-experimental experience is important in the study of human behavior. With few exceptions, however, they have tended to overlook the possible effects of early experience upon the behavior of animals. They do not specifically deny that such a problem exists; they seem rather to be either unaware of it or to neglect it. It might well be argued that they hold an implicit double standard in their disregard of such factors in the study of animal behavior.

Beach (1) points out that 50 per cent of the reports on animal experiments in the *Journal of Comparative Psychology* and its successor, the *Journal of Comparative and Physiological Psychology*, since 1930 have been made on the Norway rat, although the latter accounts for only .001 per cent of all species available for study. Since the behavior of the rat has been most intensively studied and forms the basis for the most extensive contemporary learning theories, it seems well to focus attention on the area of rat experimentation. A casual perusal of the literature will soon demonstrate that the description of the subjects follows an invariant pattern: "the subjects were N experimentally naïve rats of such and such a strain." What, then, is the essence of experimental naïveté? Explicit is the fact that the animals have never been used in an experiment. Implicit in such a description are alternate premises: either that the animals' previous experience is of no importance whatsoever, or, that it is so uniform that every *experimentally naïve* rat is a standardized little mechanism subject to exactly the same pre-experimental influences in laboratories at California, Iowa, or elsewhere.

Results of recent research suggest that it might be well to scrutinize more closely the implicit assumptions made by experimentalists who utilize rats as subjects. As yet there is not enough evidence to specify precisely what sorts of pre-experimental experiences might affect the experimental performance of rats. There is enough evidence, however, to make it worth while to examine the provocative data. The present paper is therefore addressed to the practical and theoretical suggestions posed by these findings. It will be centered around two queries: (1) Is it legitimate to assume that all *experimentally naïve* rats will display similar behavior as adults? and (2) Does the use of *experimentally naïve* rats unduly limit the findings of experimenters?

---

[1] The author is indebted to Charlotte E. Christie for surveying the recent literature and abstracting information pertinent to the pre-experimental care of animals.

THE COMPARABILITY OF EXPERIMENTALLY NAÏVE RATS

The bulk of studies published on animal experimentation in the journals of the American Psychological Association is carried in the *Journal of Experimental Psychology* and the *Journal of Comparative and Physiological Psychology*. In 1949 the former contained 18 articles describing experiments utilizing rats as subjects; the latter 38, making a total of 56 articles in this area. While scrutinizing these articles to determine the pre-experimental history of the animals used as subjects, certain interesting facts come to light.

Source. In 26 (47.2 per cent) of the 56 experiments scrutinized, *no mention was made of the source* from which the subjects were obtained. In some instances the assumption might be made that the animals were from the colonies maintained in the universities in which the studies were carried out. That this assumption need not necessarily be true is evident from the fact that in some studies in which the source was mentioned, the rats used were obtained from both the university colony and from animal supply houses as well as from other sources.

Nineteen (33.9 per cent) of the articles report that the animals were obtained from the colonies maintained by departments of psychology. In five (8.9 per cent) cases the animals were obtained from animal supply houses. Two (3.6 per cent) experiments report the use of rats obtained from nonpsychological colonies, viz., the Department of Agricultural Chemistry and the Home Economics Department. Two (3.6 per cent) other experiments utilized subjects who were partially from university colonies and in part from other sources—these in one case being procured from a local dealer, and in the other, they were wild rats from the Rodent Control Office of the City of Baltimore and the United States Public Health Service. Two (3.6 per cent) other experiments report the use of unselected stock from local dealers.

Aside from the wide variability of strains (which may or may not be reported) from such diverse sources, the factors of differential maintenance schedules, caging, and handling must be considered. When rats from nonlocal colonies are used, there is also the effect of transportation to be weighed. At the University of California the author used in experiments 130 rats from the colony maintained by the Department of Psychology. These animals were obtained when they were four weeks of age and kept until they were about four months old. During this period only one rat died. More recently the author obtained 60 animals three and a half weeks of age from an animal supply house. Three of these animals died in transit, and two more died immediately afterward. Deprivation severe enough to kill 1 of 12 animals during transportation contrasts with the much lower rate found in animals from a local colony. It seems quite plausible that such a difference in amount of deprivation or hardship might have persisting effects upon the survivors.

Maintenance schedule. Information relating to pre-experimental maintenance schedules was extremely sketchy, being mentioned in only one series of experiments. In the absence of fuller information it is necessary to rely upon personal observation. In most laboratories a constant supply of water is available during the pre-experimental life of the rat. Feeding tends to be more varied. In some laboratories the animals always have a constant supply of food available; in others they are fed once a day. The type of food varies considerably among laboratories. In some cases rats are reared upon dry mash, some on wet mash, some on kibbled biscuit, while others have a basic diet to which greens are added. It is impossible in almost all cases to determine from the data given in articles anything about the pre-experimental deprivation schedule of the animals involved, the only safe assumption being that there is considerable variability.

There is one further factor in regard to maintenance schedules which may be of importance. The responsibility for taking care of laboratory animals varies quite widely. In some instances faculty members are responsible for the care and maintenance of the animals, sometimes graduate students have the responsibility, and in other instances nonacademic employees are in charge. It is somewhat dubious to suppose that all of the aforementioned would be equally concerned to see that the animals never suffered deprivation. Within the relatively limited knowledge of the author, there are several instances in which the individual in charge of the animals neglected to feed them on schedule, with resultant deprivation. There is also the factor of accidental blockage of water bottles, resulting in unscheduled thirst for the subjects. Since there is never a mention of such factors in experimental reports, it is impossible to determine the frequency with which they occur.

Caging. Again we suffer from an almost complete lack of information. Laboratories visited by the author vary in having from one to twelve rats per cage. There is no standard cage design so that the size of the cage as well as the number of animals per cage is an uncontrolled variable. The only safe assumption is, unless the description specifically says otherwise, that the animals were reared in cages of some sort.

Handling. Once again there is no comparable information upon which comparisons of pre-experimental life might be based. In many cases when rats are obtained immediately prior to the experiment, the experimenters themselves do not know the previous history of the handling of the subjects. Aside from deliberate handling of the animals to insure greater docility, which is sometimes reported, the nature of the cages used is, in many cases, a determinant of the amount of handling received by the animals. Some cages are constructed so that droppings fall through a mesh floor and it is consequently possible to rear animals without touching them. In other cases bottomless cages rest in pans filled with shavings so that it is necessary to transfer rats to other cages while cleaning the home cages. In this fashion the animals receive a good deal of unpremeditated handling.

Aside from the quantitative amount of handling, there are not inconsiderable differences in qualitative aspects. Some experimenters dislike rats and handle them most gingerly in a nervous fashion. Others are quite at ease with animals and allow them to crawl over their shoulders, nibble their ears, and are otherwise quite permissive in their handling. In uncontrolled observations the author as well as other experimenters have often been able to tell which experimenter has handled which rats by the differences in the quality of the rats' behavior when picked up, as well as the amount of freezing, defecation, and other indices of nervousness when the animals were placed in a maze.

THE EQUATING OF EXPERIMENTAL NAÏVETÉ WITH EXPERIENTIAL NAÏVETÉ

If anyone wished to maintain that all rats have similar pre-experimental experiences, the indicated paucity of relevant information certainly leaves the burden of proof upon the protagonist of this viewpoint. To judge from the superficial amount of information available, the range of variability in pre-experimental care is considerable.

The crucial question in this regard is whether the variation in the pre-experimental life of the normal laboratory rat has any effect upon subsequent behavior in experimental situations. If not, the preceding observations are only of cursory interest; if so, they have an important bearing on the interpretations of experimental findings. Once again we are faced with a paucity of relevant information. There is, however, enough evidence regarding the effects of pre-experimental variables to suggest that the problem is a valid one.

The failure to specify the source of the subjects used in rat experiments may in certain instances conceal the fact that animals of different strains are being used. Pigmented rats have better vision than albinos and would be expected to do better in experiments in which visual cues are important. Some strains have been bred for specific sorts of abilities and will perform differently in various situations. Tryon's maze-bright and maze-dull rats were bred for performance in a tunnel maze. Krechevsky (8) found that they were responding to different types of cues. He argued that the maze-bright rats were relying upon spatial rather than visual hypotheses in problem solution.

Another unknown factor would be differential breeding in different laboratories of animals originally obtained from a common strain. How comparable are contemporary rats whose remote ancestors were of the Wistar strain? Obtaining animals from different sources also means that the previously cited variables of caging, maintenance schedules, and handling are often unknown to the experimenters themselves.

Hunt (5) gave rats experience in food deprivation when they were quite young. These rats were found to hoard significantly more pellets as adults than were nondeprived control rats and animals which had experienced an equivalent amount of deprivation later in life. This suggests that experiments involving hunger as motivation might be affected by early deprivation.

An experiment by Christie (2) indicates that early deprivation is important in an experimental test of latent learning under reversed motivation. In a modification of Kendler's (7) variation on the Spence and Lippitt (14) experiment, it was found that pre-experimental deprivation training had marked effects upon adult performance. The basic design of the experiment involved a single choice point maze with food in the right-hand goal box and water in the left goal box. Animals were trained under hunger motivation while satiated for water. They received an equal number of experiences in both sides of the maze by the application of free and forced choices. In two experiments[2] by the author with subjects which had experienced a "normal" rearing in cages with food and water constantly present, it was found that their behavior conformed to that reported by Kendler. On test trials the animals would continue to the previously rewarding right-hand side whether motivated by hunger or thirst. Two groups of rats which had been subjected to systematic pre-experimental deprivation of food and water extending between the thirtieth and ninetieth days of life displayed significantly different behavior when trained and tested under the same conditions. These animals tended to prefer the left-hand goal box containing the water bottle even though they had never been rewarded at that point.

---

[2] An investigation of the role of drive discrimination in learning under irrelevant motivation, *J. exp. Psychol.*, 1951, **41** (in press); A further examination of the nonreinforcement position in learning under irrelevant motivation (in preparation).

These experiments indicate that extensive deprivation in pre-experimental life will have effects upon adult behavior in certain sorts of experiments. The amount of deprivation necessary to produce an effect and the sort of experimental situation in which such early deprivation might be of importance are as yet unknown. The possible effects of accidental deprivation or different forms of maintenance schedules cannot be specified, although they may be important uncontrolled variables.

As far as the author knows, there is no direct evidence bearing upon the possible effects of different sorts of caging upon the behavior of rats, with the exception of one experiment by Patrick and Laughlin (**12**), who reared rats in two different sorts of cages. A control group was reared in mesh-walled cages of the sort which are generally used. An experimental group was reared in special cages with glass walls and without any shadow-producing matter in the cages. When tested for wall-seeking behavior as adults, the experimental rats showed significantly more exploratory behavior in open places. Whether rats reared in large cages would learn different things from rats reared in small cages is a moot question. Ligon (**9**) has shown that a cage mate is effective as an incentive in maze learning. Maltzman (**10**) mentions a number of as yet unpublished studies conducted in Spence's laboratory utilizing a social cage, i.e., one with cage mates, as an incentive. Possible socialization effects resulting from different numbers of rats being reared together would seem to be worthy of investigation.

The evidence on the effect of handling is, fortunately, experimentally documented. Karn and Porter (6) compared a number of pretraining experiences: (1) habituation to handling, (2) familiarity with the maze, (3) familiarity with the maze and goal orientation, and (4) habituation to being detained in an enclosure similar to the starting compartment in the maze. It was found that all of these conditions led to more rapid learning in subsequent experimentation. The first mentioned is most pertinent to the present discussion, since it has been already pointed out that differential amounts of handling might occur both deliberately and accidentally.

Although the amount of evidence is scanty, it does indicate that perhaps one of the reasons for the frequent inability of experimenters to obtain comparable results in similar experiments might be the unequal sorts of pre-experimental influences resulting from differences in source, maintenance schedules, caging, and handling. Animals from a wide variety of backgrounds are all *experimentally naïve;* this does not automatically make them *experientially naïve*, nor should it imply that their pre-experimental histories are equivalent.

To assist in the evaluation of experimental results, it is suggested that the following data relevant to pre-experimental variables be included in experimental reports of rat experiments:

1. The strain, source, and age at weaning of the animals.
2. The pre-experimental maintenance schedule.
3. The type of caging utilized.
4. The amount of handling received by the animals prior to the experiments.
5. Any accidental deprivation or other possible traumatic experience.

If these data are included in all reports, it might be possible to clarify the reasons underlying some of the differences in results reported in similar experiments. It has long been a source of wry amusement that Iowa and California rats often perform quite differently in comparable experiments. In this context differences in the pre-experimental treatment of rats are interesting. California rats are of the Tryon strain, are reared in cages from which the animals have to be removed approximately once a week while the cages are being cleaned, and are reared on dry mash. According to Professor Kenneth W. Spence (personal communication) Iowa rats are of Hall's nonemotional stock, are reared in cages which can be cleaned without handling of the rats, and are reared on pellets. In both laboratories the usual number of rats per cage is four, so one variable appears to be controlled. Whether these differences in pre-experimental life are related to differing performance as adults is a moot question. Karn and Porter (6) found that only one transfer from the living cage to another one and back once a day for six days prior to experimentation was sufficient to increase the "latent" learning ability of rats. The California rats are normally handled a dozen times or so prior to the commencement of experimental training, in contrast to two or three handling experiences of the Iowa rats. Both groups would, of course, normally receive handling the week prior to actual experimentation in order to increase docility. It is questionable whether this would suffice to equalize the effects of differential amounts of prior handling.

The author does not wish to espouse a rat phenomenology. It is suggested, however, that experimenters consider the extremes of the range of restriction of stimuli present in the pre-experimental life of the rat and reflect upon the relatively great importance which variation in these might have upon the rat's life. The more standardized the pre-experimental influences, the more valid the comparison of results obtained from different laboratories. If differing results are still found, the search for the responsible variables can be legitimately delimited to those present in the experimental situation itself.

POSSIBLE LIMITATIONS ARISING FROM THE STUDY OF EXPERIENTIALLY NAÏVE RATS

If, as the preceding discussion suggests, unintended variations in the pre-experimental life of the laboratory rat may be enough to create important differences in experimental behavior, an important question may be raised. What would be the effect of deliberately introducing drastic differences in the pre-experimental world of the rat?

The pertinence of this query arises from the fact that the average laboratory rat is reared under unique circumstances. He is confined to a relatively small area in which he has to engage in simple operations in order to satisfy his needs. The range of potential stimuli to which he is exposed is arbitrarily limited. The activities which, under more varied environmental conditions, might be learned as means of relieving physiological tensions are not developed. His world is relatively static and is not conducive to the full development of the rat's potential behavioral repertoire.

Comparable restriction of experience on the human level is extremely rare. The findings of the retardation of the children in Colvin Hollow (**13**), where they experienced a good deal more freedom than the average laboratory rat, illustrates the effect of relatively meager environmental stimulation. Feral man has not proved adaptive to the man-made mazes of civilization. If the patients of Freud had spent their infancy and adolescence in cages, it is interesting to speculate as to the sorts of personality mechanisms he might have uncovered. As it is, he has been criticized for drawing upon maladjusted, bourgeois Viennese for principles which were generalized to all mankind. If this criticism has any legitimate basis, it is difficult to justify the study of cage-reared rats in an effort to understand the behavior of rats reared in a much richer environment.

The hard-bitten rat-runner might agree with the foregoing comments, but could point out that the simple pre-experimental life of the average laboratory rat eliminates many extraneous variables and leaves the basic ones open for investigation. This is a perfectly legitimate position, although the preceding section indicates that it may be a dubious assumption in view of the uncontrolled variability that does exist. In the initial stages of theory-building simple beginnings are undoubtedly best, but it is the responsibility of the theorist to demonstrate that the principles uncovered in the study of rats with a restricted background are applicable to the behavior of rats reared under more varied environmental conditions. In the opinion of the author, this should be done before application of the principles is made to human behavior.

There is experimental evidence relative to the subsequent behavior of rats subjected to varying degrees of pre-experimental richness of environmental experience which indicates that it differs from that of rats reared under "normal," i.e., cage conditions. The experiments are few in number, but their implications cannot be neglected. In discussing them, reference will be made to quantitative and qualitative differences in learning ability. By *quantitative differences*, we mean that the behavior of the rats with richer experience differs from that of experimentally naïve rats, but only in degree. Thus, although faster or slower learning might be manifested, it would not invalidate the principles governing the behavior of the usual laboratory rat. "Qualitative differences in learning . . . " means that the behavior displayed is interpreted as involving additional principles necessary to explain the differences in performance between the two groups.

Hebb (4) reports two exploratory studies which are relevant. In the first he contrasted the behavior of two groups of rats, one subjected to "normal" cage conditions, the other to an even more restricted pre-experimental life. The latter group was blinded in infancy, but otherwise was reared under conditions identical to the first group. The animals reared under "normal" conditions were blinded at the age of three months. When both groups were five months of age they were tested in a series of experiments. It was found that rote learning tasks did not differentiate the two groups. In a task involving a change of response given at eight months of age, it was found, however, that every late-blinded animal did better than every early-blinded animal. Hebb interpreted this as indicating the permanent effect of early experience upon problem-solving ability at maturity.

The finding that animals subjected to meager early experiences did more poorly than animals subjected to "normal" rearing suggested a further exploratory study. What would happen if "normal" rats were contrasted with a group of rats who had richer early environmental experience? Hebb took seven animals home and reared them as pets.

During this period of time they experienced a good deal of time outside their cages, running about the house. When these pets were compared with cage-reared rats in a total of 21 test situations, it was found that the pets all scored in the top third of the total distribution for cage-reared and pets (25 cage-reared animals were used). Inasmuch as Hebb does not describe the testing situations, it is impossible to determine whether there were qualitative as well as quantitative differences in their problem-solving behavior.

An experiment by Christie (2) indicates that qualitative differences do appear, at least in certain experimental situations. In a modification of Kendler's (7) experiment previously mentioned, a group of rats which had been given extra-maze exploratory experience was tested. The exploration training was given under satiation on an elevated runway and an open table top. Food and water were present on the table, and the animals received 20 exploratory experiences of 24 hours' duration between the ages of one and three months. When these animals were tested under conditions identical to those previously alluded to, it was found that they displayed behavior significantly different from that of cage-reared animals which also had never experienced deprivation. The early explorers did go to the water bottle and drink when thirsty, as contrasted with cage-reared rats which continued going to the goal box containing food where they had been previously reinforced. Richer early experience in this case quite definitely led to qualitatively different behavior.

These findings led to another series of experiments (not yet prepared for publication) designed to determine what effects early exploration would have upon the behavior of rats in place-learning experiments. Rats which were given extensive experience upon changing configurations of tables connected with runways were tested, as were cage-reared controls. Food and water were always present on one of the tables and in the cages. The exploratory training extended through the second and third months of the rats' lives.

The first experiment required the animals to travel from a starting point fixed in space to a goal which was also fixed in space. The route between was a multiple T-maze with nine choice points. The maze was shifted on every trial so that the animals never traversed the same path from the starting point to the goal. Under these conditions it was found that both groups of animals displayed indications of place learning but that the early explorers learned more rapidly, so the differences were quantitative rather than qualitative.

In a second experiment the location of the goal was fixed, but the starting point was shifted in space. The route to the goal also varied from trial to trial in this experiment. After a training series the animals were placed upon a new starting point and had an option of choosing from a number of radiating paths, one of which led directly to the goal. The cage-reared animals made chance responses to the correct path, whereas the early explorers chose the correct path at a significant level. In this situation we find qualitatively different behavior differentiating the animals on the basis of their early experience.

In a third, still more complex situation, neither group gave significant evidence of having learned the problem. Thus, in situations of varying difficulty, it was found that in the simplest the early explorers displayed a quantitative superiority, in a more complex situation a qualitative superiority, and in a very difficult situation no difference in performance was demonstrable.

Interesting differences in the rats' behavior during all the testing were observed, these differences bearing no known relationship to learning criteria. At the inception of the experiments, all rats were given four practice trials on running down runways to food. The early explorers were undisturbed by this training. The cage-reared animals behaved as might be expected. They were quite excited, and there was a high incidence of freezing and defecation. They would not voluntarily advance down the elevated runway but had to be coaxed.

For all three experiments the median time per trial was plotted against the number of trials. Both groups show a typical performance curve in running-time for all three experiments—starting slowly and then reaching a plateau. This plateau was reached after an equivalent number of trials for both groups. Of special interest is the fact that the running-speed of the early explorers was consistently twice as fast as that of the cage-reared animals during all three experiments. At the completion of the last experiment the cage-reared animals had had a month's daily experience on elevated mazes; yet the difference in running time was still significantly faster for the early explorers. Quite extensive adult experience on elevated mazes on the part of the cage-reared rats was not sufficient to compensate for the early exploratory experiences received by the experimental group.

This difference in gross behavior of the two groups was so great that observers who had never seen rats run a maze could readily differentiate between the two groups on the basis of the manner in which they ran the maze. The early explorers were not only much faster but ran the maze in what might be described as a "slap-dash" manner. When approaching choice points (all of which were new to them on every trial), they displayed little hesitancy and would often make a choice without slowing down. An indication of this abandon is illustrated by their behavior when they chose blinds. They would occasionally be running so rapidly that they would be unable to stop at the end of the blind alley and would skid over the end of the maze and slide half-way down the support to the floor before they could stop. They would then climb back up to the maze and start running again, with no observable effects on their subsequent behavior. The cage-reared rats never displayed such behavior. No cage-reared rat ever ran off the end of a maze. Cage-reared rats also displayed different behavior at choice points, since they tended to be much more hesitant with a good deal of freezing and "VTEing."

If further experimental work confirms the initial finding—namely, qualitatively superior behavior in problem-solving situations when rats with a wide degree of pre-experimental experience are used as subjects—certain implications follow. A greater range of observed behavior leads to a broader empirical basis for theoretical conceptualization. Just as the factor analyst is limited in his findings by a narrow range of behavioral correlations, so is an animal theorist limited by the narrow range of behavior exhibited by cage-reared rats.

A further area which could profitably be investigated would be that of various genetic hypotheses in which rats are used as experimental subjects. A good deal of the emphasis in personality theory has been on developmental aspects of learning. These have been, as Gibson (3) points out, relatively neglected in the area of animal experimentation, Hebb (4) being a refreshing exception. Many possible permutations of early deprivation and types of exploration and learning are possible which woud be of considerable potential value in clarifying various relationships of early experience to later behavior.

The continued equating of *experimental naïveté* with *experiential naïveté* means quite bluntly, in the author's opinion, that animal experimenters are not studying the principles of behavior, nor the principles of rat behavior, but the behavior of cage-reared rats. For many purposes this is quite adequate; for a more generalized theory of behavior it seems to fall short of asking what seems to be an extremely important question: What are the effects of earlier experience upon subsequent behavior?

SUMMARY

A survey of recent articles on rat experiments reveals that very meager information is available on the pre-experimental life of the *experimentally naïve* subjects. As far as can be determined, a considerable amount of variability in treatment occurs in various laboratories. Experiments were cited which indicate that these variations might have an appreciable effect upon the behavior of adult rats and might underlie some of the contradictory findings reported in the literature. Suggestions for additional data on the pre-experimental care of rats were made so that comparisons between experiments might be more validly made

It was further argued that the exclusive use of cage-reared rats might unduly limit the body of behavioral data, thus having a restrictive influence upon the theoretical inferences made from analyses of the data. Recent experiments were reviewed which indicate that rats with early exploratory training exhibit as adults qualitatively different behavior from that of rats without this early training. Further research designed to ascertain the effects of differential pre-experimental training is needed.

BIBLIOGRAPHY

1. Beach, F. A. The Snark was a Boojum. *Amer. Psychologist*, 1950, **5**, 115–124.
2. Christie, R. The effect of some early experiences in the latent learning of adult rats. (To be published.)
3. Gibson, J. J. The implications of learning theory for social psychology. In J. G. Miller (Ed.), *Experiments in social process*. New York: McGraw-Hill, 1950.
4. Hebb, D. O. *Organization of behavior*. New York: Wiley, 1949.
5. Hunt, J. McV. The effects of infant feeding frustration upon adult hoarding. *J. abnorm. soc. Psychol.*, 1941, **36**, 338–360.
6. Karn, H. W., & Porter, J. M., Jr. The effects of certain pre-training procedures upon maze performance and their significance for the concept of latent learning. *J. exp. Psychol.*, 1946, **36**, 461–469.
7. Kendler, H. H. An investigation of latent learning in a T-maze. *J. comp. physiol. Psychol.*, 1947, **40**, 265–270.
8. Krechevsky, I. Hereditary nature of 'hypotheses.' *J. comp. Psychol.*, 1933, **16**, 99–116.
9. Ligon, E. M. A comparative study of certain incentives in the learning of the white rat. *Comp. Psychol. Monogr.*, 1929, **6**, 1–95.
10. Maltzman, I. M. An interpretation of learning under an irrelevant need. *Psychol. Rev.*, 1950, **57**, 181–187.
11. Miller, D. R., & Hutt, M. L. Value interiorization and personality development. *J. soc. Issues.*, 1949, **5**, No. 4, 2–30.
12. Patrick, J. R., & Laughlin, R. M. Is the wall-seeking tendency in the white rat an instinct? *J. genet. Psychol.*, 1934, **44**, 378–389.
13. Sherman, M., & Henry, T. R. *Hollow folk*. New York: Crowell, 1933.
14. Spence, K. W., & Lippitt, R. An experimental test of the sign-gestalt theory of trial and error learning. *J. exp. Psychol.*, 1946, **36**, 491–502.

Received August 28, 1950.

THE USE OF EXTENDED CONTROL-GROUP DESIGNS IN HUMAN RELATIONS STUDIES

RALPH R. CANTER, JR.
University of California

Most psychologists who have conducted experiments in complex social situations wherein a single control group was used, have wondered privately and publicly as to the exact genesis of an observed change in their experimental groups. Several alternative explanations usually seem to be required to satisfy the wonderment. These are quite likely offered because of the psychologist's awareness that complex variables have been continuously operating and could not be cut down to laboratory size without distortion of an unknown magnitude.

While further experiments upon hypotheses derived from obtained changes are certainly desirable in order to specify the ways in which these variables are operating, the complexity and practicality of the situation, e.g., in industry work groups, is often such that subsequent experiments cannot be carried out.

What is needed are more designs and methods which will yield estimates of the operations of several variables within one experimental period, since subjects drawn from a complex societal group often reach a satiation point with one experiment.

A pertinent example of such problems is found in the rather poorly defined area of human relations. Experiments in this area primarily deal with some obfuscated relationships among persons interacting according to some psychosocial system, e.g., superiors-subordinates, members-group, members-nonmembers, husbands-wives, professionals-nonprofessionals, participants-reactivists, elected leaders–appointed leaders, etc.

Usually most psychological experiments in this area are set up to effect some kind of behavior change with a view to determining the producers of such a change. Unfortunately, these experiments often yield responses from the subjects which are difficult to relate to antecedents. For example, an experiment on factory absenteeism may have the hypothesis that a "high" rate of absenteeism is associated with intragroup conflict. All groups are compared on absenteeism rates, and those with similar "high" rates are studied further. Accordingly, systematic surveys and interviews are conducted to locate two groups—one experimental and one control—equally matched on mean intragroup conflict ratings (assigned scores to each individual on a given scale).

This constitutes the pretest. Then, a series of group meetings, led by a skilled, permissive, "democratic" leader, directed toward reduction of conflict, is used as the independent variable in one group, the other group remaining as a control. The posttest involves a repetition of the pretest, measuring the extent to which changes have taken place in conflict scores assigned by raters, and also observing the absenteeism rate. Let us suppose that little difference in change is found between experimental and control groups' results on either count and that the circumstances are such that it is not possible to modify or replicate the experiment on the same groups.

Granting that the matching variables also constitute the pretest, the fact remains that in this hypothetical experiment the usually effective techniques, for reducing conflict at least, failed to work. The experimenter may reason along several different pathways and not meet with an hypothesis that will meaningfully account for the results.

Recently, Solomon (8) has made a suggestion for the use of additional control groups in certain kinds of experimental studies. This extension is particularly applicable to studies in the area of human relations, especially where only a single experiment is possible.

He proposed an extension of the traditional single control-group design that allows for estimates of some of the variables arising in the experimental situation which are not easily noted in a single-control design. This involves the selection of a third group (called the second control group, C2) which does not receive the pretest (being assigned the mean of the E and C1 values), but which receives the same treatment as the experimental group and the posttest. A value *I* (interaction effect) is obtained by adding the amount of change taking place in the C1 and C2 groups, and subtracting this from the amount of change obtained in the E group. Solomon reports that as yet there is no means for reporting significance, only magnitude.

Schematically, his design is as follows:

	Experimental Group	*Control Group 1*	*Control Group 2*
Pretest	Yes	Yes	No
Training	Yes	No	Yes
Posttest	Yes	Yes	Yes

It should be noted that he also provides a four-group design, wherein a C3 group is given the posttest but neither the pretest nor the experimental treatment.

The essential reasoning behind his extension is that there are effects of the pretest combining with the independent variable and subsequently interacting with the posttest. These interaction effects seem to arise from attentional and/or perceptual shifts due to the pretest. Using training as an example, he presented several assumptions: (1) that the improvement of C2 over its assinged mean is due to training alone; (2) that the improvement of C1 is due to the facilitating effects of the pretest alone; and (3) that the possible failure of the improvement of the experimental group to be equal to the sum of the improvement of C1 and C2 is a reflection of the interaction effects of the pretests. Now, the value *I* may be positive or negative in sign. If positive, then it would appear that the pretest somehow had a beneficial effect on the results of the training procedures. If negative, then the pretest may have developed some kind of resistance or interference: emotional, attitudinal, perceptual, or general orientational "sets."

The value of *I* may well be related in some fashion to Harlow's (**6**) learning sets, i.e., readiness, or learning to learn; to Bateson's (**1**) concept of "deutero-learning"; to Cronbach's "test response sets" (**5**); and to the somewhat neglected "principle of expectancy" as shown in such studies as those of Cowles and Nissen (**4**) and Tinklepaugh (**9**). In the latter's experiment monkeys rejected lettuce when they were apparently "expecting" a banana, though they normally accepted lettuce. In our hypothetical experiment perhaps the experimental group might reject participation in discussions if they were led to expect lectures in "how to get along with people," etc. At any rate, the value of *I* may possibly be a quantifiable expression of such factors as learning contexts, the operation of expectations, sets, and so on.

Applied to our hypothetical problem, this design would provide an inference as to whether the pretest in some fashion operated unfavorably upon the E group meetings. For example, the nature of the interview might have led the participants to anticipate that certain points would be brought out in the discussion in the same way they were in the interview. When these did not materialize in a recognizable form, the participants might easily have cast out the whole proceedings as a waste of time. Since these factors could not hold for the C2 group, this group might engage very adequately with the points brought out and find means for resolving the group difficulties. Further, absenteeism might well be reduced in this group (providing the hypothesis were tenable), the group members accepting nonconflict roles, being motivated to maintain a conflict-free group, and so on.

Within our hypothetical human relations experiment, one difficulty is pointed up in a two-control-group design, How would we select our C2 group? The method deliberately chosen was that of systematically surveying groups and finding two groups with equal absenteeism rates to be further matched on the variable of intragroup conflict, this constituting the pretest. But C2 is not supposed to have the pretest. It would seem that there are several ways at hand for obtaining this group.

1. One involves the judgment of a person or persons sufficiently familiar with the various groups to be competent to designate the C2 group. But in human relations studies, this presents a very difficult assumption due to the nature of the variables. The superintendent of a factory may be able to show absenteeism records, but he cannot be expected to rate conflict on a common basis. His communication usually depends upon his subordinates, who are often motivated in ways at variance from the existing situations, i.e., the subordinates may not like others to know of the existence of intragroup conflict.
2. A second method for obtaining the C2 group would involve random or stratified sampling, drawing persons from various groups. These persons should not be from groups having had the pretest. However, this would mean individuals who did not know each other or who knew each other only slightly. The conflict variable could hardly seem relevant here. Only an assumption that they would be in conflict were they to be formed in a work group would make this selection tenable (though the writer often feels that this assumption would be entirely in order!).
3. A third way would be to select persons matched on any number of variables with the E and C1 members. For example, age, education, sex, length of service, ratings, occupational position, absenteeism record, etc. However, these variables may in no way be correlated with social interaction producing conflict. This could be determined, of course, by pre-experiment studies after the E and C1 groups have been chosen. But the likelihood of other than chance relationships seems small.
4. A fourth method, a variant of the way in which the E and C1 groups were obtained, would be to find the groups with equivalent absenteeism records and attempt to match the group's histories with those of the E and C1 groups. Many different sources would be utilized: the opinions of all persons familiar in any respect with each group, leadership histories, exit-interview results, production variability, medical records, variability in size, long-term reputation among other groups, and anything else that might be relevant. However, these would provide equivalent problems to those raised in the third method—pre-experimental studies would probably be required. This method does have, however, the advantage in that these variables seem to have greater face validity.
5. A fifth way would be to use free-floating trained observers who, without using the formal pretest, might uncover a likely-looking group. But conflict has an elusive quality when being observed. The shop gets cleaned up in a hurry when officialdom is suspected.

Thus, for human relations studies wherein the variables are quite complex, this design has an immediate difficulty inherent in it. Nevertheless, Solomon's suggestion provides a generalized design urgently needed for such studies and should engender a great deal of further research.

AN EXAMPLE OF THE USE OF THE TWO–CONTROL-GROUP DESIGN

The writer elsewhere (2, 3) has reported in detail a study of the development and evaluation, by tests and other measures, of a human relations training course for executives and supervisors within the home office of three insurance companies. As part of the evaluation, a C2 group was used. While not of the order of the hypothetical experiment discussed above, certain aspects are relevant.

As in most industrial studies certain problems are inherent in the working situation. The relevance of any demands upon the working time of employees is properly of concern to executives, for example. In this study, it was thought that any course involving "psychology" and "human relations" should be fully understood by the executives prior to being tried out experimentally with their supervisors. Hence, a series of ten sessions involving about 45 executives was held. Then the experimental program was launched. All supervisors from one department formed the E group, and all supervisors from two other departments formed the C2 group, these latter departments, combined, equalling the E department in size. All three departments performed identical work, faced the same kinds of problems, and were very comparable in many other respects.

While the general design was satisfactory, there was one important problem. The E and C1 groups were both selected from departments whose executives had attended the executive training series. Ideally, none of the executives would have received training in the control departments. As noted above, it was considered important that the executives hold positive attitudes toward the training. To choose an E or C1 or C2 group whose executives were not familiar with what their subordinates were being requested to do was considered dangerous to the experimental study. Negative attitudes could well have had vitiating effects upon measured performance. Hence, improvement in the performance of the C1 group could have taken place through the communication of information and "democratic attitudes" from the trained executives to their untrained supervisors. This would then result in an underestimation of the improvement of the E group over the C1 group.

A C2 group was then selected. Faced with the same problem, that the executives had to be informed, the experimenter enrolled only supervisors whose superiors had attended the preliminary series. (The course was apparently quite popular—many more requests were received for enrollment than could be granted—especially from departments whose executives had attended the first series.) Again, ideally none of these supervisor-executives would have been trained. Since the problems involved in the selection of this C2 group had not occurred to the writer at that time, he accepted one supervisor from each department, with the exception of three large departments (100 or more employees) wherein two each were accepted. Also, since all members of the C2 group were from different departments, the design was further weakened. This general selection method corresponds most closely with the second selection method discussed above.

The final *N* was 18 in each of the E, C1, and C2 groups. The groups were found to be fairly closely matched on means of age, education, mental alertness test scores, and years of service, and approximately equal in proportions of males and females.

A total of 12 separate scores was obtained from the measurement battery used. The value of *I* has been determined for 9 of the measures. To reduce discussion, Table I has been prepared showing the obtained *I* values and *t* values.[1] (A full description of the measures can be found in reference **2**.)

From these results it seems that differential effects of the pretest

TABLE I

MAGNITUDE OF OBTAINED *I* VALUES BETWEEN PRETEST AND POSTTEST SCORES AND OBTAINED *t* RATIOS BETWEEN PREDICTED AND OBTAINED POSTTEST SCORES

Measure	*I**	*t*†
General Facts and Principles	−2.98	0.0001
How Supervise? (total)	−0.23	.0411
How Supervise?		
I. Supervisory Practices	+0.67	.2039
II. Company Policies	+0.72	.0465
III. Supervisory Opinions	−1.61	.0760
General Logical Reasoning	+0.05	.1805
Social Judgment Test	−2.03	.3050
Test of Ability to Estimate Group Opinion:		
I. Department Error	+7.75	.0270
II. Department Bias	−7.30	0.0241

* *I* = interaction effect of pretest and training on posttest scores.

† *t* = ratio of standard error of obtained differences between predicted and final posttest scores to its mean difference.

---

[1] Student's *t* was used, being obtained from a comparison of predicted and obtained means of the E group. Each predicted mean was derived from the regression of the C1 posttest scores on its pretest scores according to a covariance design presented by Peters and Van Voorhis (7, pp. 361–364). The C2 group did not enter into this analysis.

can be inferred in several measures. A brief discussion of a few of the individual tests will assist in gaining some insight into the nature of these effects.

The General Facts test is a short test of general psychological knowledge. Few of the trainees had any information about psychology. But they appeared to hold some fairly definite opinions about what it was. Hence, the structuring presumably given by the pretest could be assumed as establishing a "set" which, interacting with their preconceptions, might have served to make the course somewhat confusing to them. They gained significantly over their controls, but perhaps not as much as possible.

The "Supervisory Opinion" section of the File-Remmers How Supervise? test was quite similar in content to the General Facts test. The same result may have held here. The other two sections of this test seemed to reveal that the pretest established a facilitating orientation for the material obtained in the training, while the C2 subjects did not have this experience.

The Social Judgment test was acknowledged when taken the first time by the trainees as being quite difficult and somewhat ambiguous to them. Very few complained about it on the posttest, however. It did not reveal any significant improvement, as shown by the t value. Yet a sizeable I was obtained.

Two rather surprising findings occurred in the two subtests of the *Test of Ability to Estimate Group Opinion.* Simply stated, the error score represented how many percentage points the supervisors were in error in estimating the percentage of persons in their departments holding certain opinions (morale scores) about working conditions, company policies, community, their jobs, their supervisors, etc. The bias score represented the direction of their errors, i.e., underestimation or overestimation of the percentage. Apparently the E supervisors were more correctly oriented as a result of the pretest as far as error was concerned, but the over-all bias (which was a sizeable underestimation of morale) did not improve as much as possible. Again, this was seemingly a function of the pretest effects. These results lead to an hypothesis that experience in judging morale has the effect of making the judge at once more sensitive to the group's opinions, but also more "defensive" as to the meaning of those opinions when he himself is involved. In this connection the pretest could point up the areas where the judge should look for morale, but also put him "on the spot" so that he becomes further biased in interpreting the meaning. At least in these results an apparently rather strong inhibiting effect was present.

The major result of this brief and minor study has been to point out that this is a new method which can be of value in providing a broader theory base in the study of human relations training. More extensive and better controlled conditions can result from planning studies which give some indications of the effects of the measures themselves upon the subjects' responses. A great deal of research must be done, however, upon the value of I under different conditions, particularly in regard to estimation of significance. It is hoped that this report will serve as a stimulus in this direction.

SUMMARY

This paper discussed the need for and the use of multiple control group experimental designs in one area of current research activity, namely, the study of human relations training. Certain problems in the selection of the control groups were noted, and a brief report of the use of the two–control-group design was presented. Further studies of the use of this interesting and important method were urged.

BIBLIOGRAPHY

1. BATESON, G. Social planning and the concept of "deutero-learning." In T. M. Newcomb and E. H. Hartley (Eds.), *Readings in social psychology.* New York: Holt, 1947. Pp. 121–128.
2. CANTER, R. R., JR. *An experimental study of a human relations training program.* Unpublished Doctor's dissertation, Ohio State Univ., 1949.
3. CANTER, R. R., JR. An experimental study of a human relations training program. *J. appl. Psychol.*, 1951, **35**, 38–45.
4. COWLES, J. T., & NISSEN, H. W. Reward-expectancy in delayed responses of chimpanzees. *J. comp. Psychol.*, 1937, **24**, 345–358.
5. CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt*, 1946, **6**, 475–494.
6. HARLOW, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 50–65.
7. PETERS, C. C., & VAN VOORHIS, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940.
8. SOLOMON, R. L. An extension of the control group design. *Psychol. Bull.*, 1949, **46**, 137–150.
9. TINKLEPAUGH, O. L. An experimental study of representative factors in monkeys. *J. comp. Psychol.*, 1928, **8**, 197–236.

Received October 3, 1950.

THE COMPARISON OF SERIES OF PROPORTIONS: THE PROBLEM AND A NEW SOLUTION

R. C. DAVIS
Indiana University

I. The Problem

Experimental data are often produced in the form of proportions of responses of stronger-weaker calls, yes-no reports or successes and failures to a succession of stimulus magnitudes. The point of an experiment will then usually depend on a comparison of the trend in one series of proportions with that in another. The classic example of data of this sort is, of course, that provided by a series of judgments in which a subject has tried to discriminate the intensity or some other characteristic of the stimulus.

Proper mathematical treatment of such data was a problem studied by Fechner and by psychophysicists for the half-century following him, and the most thoroughgoing solution is usually thought to be the famous process developed by Müller and Urban about fifty years ago. At about the same time came the basically different proposal of Spearman: that of computing the mean of a distribution constituted by the differences between successive proportions. Since that time no basic innovations seem to have been adopted.

The viewpoint of the founders of psychophysics was evidently somewhat different from that of the modern experimenter. The major concern with them was to calculate a set of parameters which would give "the best" description of a set of data. There is also the suspicion that they were sometimes more interested in the problem as a mathematical puzzle rather than as a research tool, and carried the refinements beyond the limits actual data would permit.

From the point of view of the present-day experimenter, the classical methods need re-examination with whatever help can be obtained from modern statistical developments. A process which gives the best description for one purpose may not serve so well for another.

A particular requirement of the present-day experimenter is that he needs to know the sampling error and distributions of the parameters he is comparing. Bruner and Postman (**3**) have discussed the importance of the error figure in relation to the study of time error in psychophysical judgments. With estimates of the sampling errors the experimenter is able to decide with some confidence whether to accept or reject the null hypothesis for the differences which appear in the parameters from one situation to another.

II. Previous Proposals

The classical methods were developed at a time when less consideration was given to the need for a probable error.

With the development of general statistical practice at the beginning of the century a number of psychophysicists set themselves the task of deriving an expression for the standard error of the traditional constants, originally in order to decide which method of calculating the parameters is best. Three approaches were used. Urban (**12**) took as his starting point the deviations of experimental values from the ogive which had been computed as the best fit for them. Thompson (**11**) started from the standard error of proportions: $\sqrt{pq/N}$. Boring (**1**) and Culler (**4**), on the other hand, followed the line of reasoning that had been adopted in common statistical techniques and attempted to derive the standard errors from the standard deviation of the whole distribution. Culler's solutions have evidently been regarded as final, for they are repeated in the standard works on psychophysics and used in some important researches (**8**, **14**). Nevertheless, as will be shown, they do not bear up under examination.

Culler made two attempts at applying the formula

$$SE_M = \frac{SD}{\sqrt{N}}$$

(or its equivalent) to the case in hand. Reasoning chiefly by analogy, Culler attempted to make this transfer in his first paper, then realized that the result had some peculiar implications and decided that the application was incorrect. In his second attempt he used a further analogy to arrive at a much more cumbersome expression. Applying this to an illustrative set of data, he found that it gave nearly the same numerical result as the simpler formula if some, but not too many, of the data were discarded. This leaves the support for the simple first formula very slender indeed, but it is the one which is recommended, without very much reservation, by Guilford (**8**) and by Woodworth (**14**).[1]

It is essential, therefore, to look into the validity of both formulas. The Müller-Urban process, of course, can yield a measure of the *SD* of the curve fitted to the data. If this could be taken as having the same properties as an "observed" *SD*, and if a value of *N* could be obtained, then the conventional formula could be used. Culler's first proposal was that the psychometric curve should be regarded as a cumulative distribution showing the actual number of thresholds below any given stimulus magnitude. It seemed, therefore, that *N* would be the total number of thresholds distributed and hence equal to the number of judgments at one stimulus intensity. (Somewhat similar reasoning had been used earlier by Boring, and rather cautiously, by Brown and Thompson [**2**].)

---

[1] In the summary of his second paper Culler's statement about the simple formula is much stronger than that in the body of the article.

This formula has some strange consequences, some of which Culler recognized in his second paper. Since *N* is in no way affected by the number of comparison stimuli used, the experimenter would attain the greatest exactitude in his measurements if he used as few as possible, and concentrated all observations on them. If he makes observations on any other comparison stimuli, he might as well discard them, since according to the formula they will not make his results a bit more precise.

In an attempt to get around this absurdity Culler proposed his second formula. In this, the basic procedure is that *N* should be determined by applying the Müller-Urban weight to each proportion of greater judgments and summing the results. Since no proof is given that this will yield a proper value of *N*, the procedure is hard to criticize. But it does have the following implications:

1. The numerator (*SD*) and denominator (*N*) of the expression for error are made to depend on the same figures.
2. The experimenter would obtain greatest precision in his mean and *SD* if he concentrated all his observations as close as possible to the 50 per cent point, where Müller-Urban weights are greatest.
3. Except in a special case, operations on the judgments "greater" and judgments "less" will give different values for *N*, even with only two categories.
4. If Müller-Urban weights are not used in calculating the median and *SD*, presumably they should not be used in computing *N* for the error formula. Then *N* becomes the total number of judgments for all stimuli, apparently reducing the error, so that computation without Müller-Urban weights gives a more precise value than with them.

Application of the formula to a set of data convinced Culler that a further empirically derived correction was necessary if one wants better than 50 per cent accuracy in the value of the standard error, when the data are not homogeneous. The above deductions strongly suggest that there is something more seriously wrong with the formula.

The meaning of *N* in statistical practice seems to be the number of independent observations that have been made on some variable. In other words, the experimenter decides on N and observes the value of x that many times. With constant stimuli the situation is otherwise: the experimenter sets a value of x and observes the number of times one of several alternatives occurs. In the first case summing the frequencies in all intervals gives the number of magnitudes observed, and also the number of observations, since each is observed once. In the second case summing will give the number of magnitudes observed but not the number of observations made. Here seems to be the basic difficulty which prevents the analogical use of the "total greater judgments" as N of the error formula.

It is a tempting alternative to identify N of the formula with the total number of reports obtained from *O* (the implication of Culler's second formula). These are observations of a sort, to be sure, but for N we need "observations of the magnitude of the x variable" (measurements of transition points in this case), and this *O*'s reports are not. Measurement of an x value assigns a certain number to it, with the implication that it lies within the interval that number represents. A single report from *O* merely locates a transition point as above or below a certain figure—places it, so to speak, in a class with only one limiting boundary. So the total of *O*'s reports cannot be equivalent to the N of the formula.

A recent proposal by Finney (**7**) apparently has a similar starting point, but the basis is the presumed variance of the transformed p-values, rather than the characteristics of the fitted curve. (Finney's paper is primarily concerned with a method of weighting the data and fitting the curve, but only his method for finding error variance is in point here.) To Finney it seems obvious that the variance of the mean transformed ordinate (mean gamma-value) is $1/\Sigma nw$, where n is the number of observations made at each stimulus value, and w the appropriate weight. Likewise, the variance of the slope of the fitted line (the reciprocal of the standard deviation) is said to be $1/\Sigma x^2$, where x is deviation from the mean stimulus. By using these two expressions and the fitted equation, he derives an expression for the variance of the stimulus value corresponding to a 50 per cent division of judgments.

But the two expressions can hardly be correct for the reasons already set forth: there is a confusion in the meaning of "σ" and "n." The implications of these formulas are, therefore, very strange. For the mean gamma-value the variance is said to be determined only by the weighted n's, with no effect of the number of stimuli. Hence the greatest precision would be obtained by concentrating all observations on one stimulus in the region of greatest w. For the slope, which would be used as a measure of discrimination, the variance is determined entirely by the stimulus values, with the number of observations and *S*'s responses making no difference. The variance is determined before the experiment begins, and an experimenter would do as well to make only one observation per stimulus. Clearly these expressions are faulty.

The other lines of attack on the problem, which actually came earlier, have a more explicit mathematical basis. Urban (**12**) carried out a derivation based on the deviations from a fitted line, but found results he could hardly believe when he applied it to an actual problem. In spite of criticism by Thompson (**11**) and Culler (**6**), at last report he seemed to be standing pat nevertheless (**13**). (The literature is very controversial. Irwin [**9**] surveys some of the arguments.) Thompson's own derivation (**11**) was an attempt to find the errors of the constants from the theoretical errors of the various proportions experimentally observed. The result, when applied to Urban's data, gave a more credible answer differing from the "empirical" figure by a factor of about 2. The Thompson formula is extremely complex, and the calculations it requires are forbidding. (Urban thinks them spoiled by required approximations, though this is not clear.) The formula, it seems, has never been used by anyone, but it needs to be considered as a possibility.

Thompson took it as obvious that the correct expression for the *PE* of the median or mean (fitted) is

$$w_a{}^2 = \sum_1^n \left(\frac{da}{dp_k}\right)^2 w_{p_k}{}^2$$

where

$w_{p_k} = .6745\ \sqrt{pq/N}$,
a = the median,
p_k = the proportion of "greater" judgments at each stimulus value as shown by the fitted ogive.

The expression in the parentheses is then evaluated by differentiating the "Normal Equation" (used in the Müller-Urban process) which relates a to p. The rationale of the basic equation evidently involves these three ideas:

1. The *PE* (w_{p_k}) of any point (p_k) on a fitted ogive can be secured by the usual formula for the probable error of a proportion. This, of course, would give a probable error in terms of the ordinate of the ogive and would, if correct, indi-

cate that in the long run half the values of p_k would lie within the limits $p_k + w_{p_k}$, and $p_k - w_{p_k}$.

2. Since it is desired to have the *PE* of the abscissa value a in terms of the abscissa scale, it is supposed that the ordinate *PE* can be converted into a *PE* on the abscissa if it is multiplied by an appropriate value (da/dp).
3. The *PE* of a median based on a series of proportions is taken to be the square root of the sum of the squares of the converted errors of all these proportions.

Now the first of these propositions seems as though it ought to have some proof (the requirement of independence may not be fulfilled), and the other two seem erroneous. In the case of assumption 2, it needs to be remembered that w_{p_k} is not a single proportion but an interval on the scale, containing a range of proportions. For each one included it would seem that a different multiplier would need to be used. If that were done, a different value for w_a would result, and the distribution it represents would be radically changed.

If one looks at Thompson's formula carefully it becomes apparent that an experimenter would not merely waste his time by adding more stimulus values; he would apparently be making his work less and less precise, even though he should increase the number of observations proportionately. Surely it must be that assumption 3 is incorrect. The idea seems to have been transferred from another situation not very closely analogous.

Thus all three attacks on the problem have bogged down. There is even no reason to hope that one or the other of the proposals will give an approximate answer: if we do not know what mark we are shooting at it would be foolish to pretend we have nearly hit it. Actually, an experimenter who followed the implications of any of the proposed formulas would be led to absurdities.

III. A Practical Solution

If one sets the problem in a way slightly different from that of the classical psychophysicists, an easy solution is available. In the "constant process" there is first a transformation of p-values to corresponding gamma values (or *SD* values) by means of the normal probability table. (In the full process they are thereupon weighted.) At this point we may picture a graph on which the points are plotted, the ordinate representing gamma or *SD* values, and the abscissa stimulus values. According to the constant process, then a line of "best fit" is secured by the method of least squares, a line whose mean and *SD* in stimulus terms are taken as the psychophysical parameters. In this application of the method of least squares it is the practice to minimize the deviations (weighted) from the fitted line on the *y*-axis. For certain purposes this seems a reasonable value to minimize, since the variations in the experiment are primarily in the subject's responses, which yield the gamma values. This process gives a line, it will be noticed, which is identical with one of the well-known regression lines, though the equations given for securing its constants look more elaborate. This regression equation is of the form

$$\bar{y} = r\frac{\sigma_y}{\sigma_x}(x - M_x) + M_y,$$

the form which is appropriate for predicting the most probable *y* (gamma—*SD*—and eventually *p*-value) from a particular *x* or stimulus value. Now the data in hand in an experiment are naturally a set of *p*-values. So in order to use the regression equation above, it is necessary to turn it around algebraically, making *x* the value to be calculated from a specified *y*-value. The intention evidently is to find a stimulus value which will most probably give a certain proportion of judgments in one category if it is presented a great many times.[2]

But one may set out with a different requirement in mind: to find the most probable value of the stimulus which yields a certain proportion of judgments. When the problem is stated so, of course the other regression equation is called for:

$$\bar{x} = r\frac{\sigma_x}{\sigma_y}(Y - M_y) + M_x.$$

This equation is derived by fitting a line as described, but minimizing the squared deviations on *x* instead of on *y*. This seems an unreasonable thing to do, since *y* is undoubtedly the major source of variation. Yet it is the operation called for to find the most probable (or, at any rate, least-squares) value of the stimulus yielding a given proportion of judgments. Using it, one can find the most probable stimulus value for the mean of the gamma-values, and the most probable stimuli for the mean of the gamma-values plus and minus 1 *SD*. It is not necessary to use this equation in reverse; one enters it with a *y*-value and solves for *x*. This method of procedure we may call the stimulus regression method.

The numbers produced by the two processes will not be the same;

---

[2] This use of the regression in reverse makes the expression for the second constant, the mean, very complex. It is derived by first getting an expression for an ordinate value called *c*, then multiplying it by the *SD* (or dividing by the *h*) secured from the first equation. The method proposed in the following avoids this complication by using a regression in the normal direction.

they will approach equality as the correlation increases. (Stimulus regression values of *CE* and σ are equal to the conventional values multiplied by r^2.) Since the r in most sets of psychophysical data is very high, the differences between the values will be small. Although one might argue that the new way of framing the problem is better than the old even in theory, the point is scarcely worth debating in most cases because the values will be so close together anyway. (See Table I.)

The great advantage of the second approach is quite plain: the standard error of estimate will provide a measure of sampling error for obtained values. This standard error is derived very simply from r and σ_x, and so avoids the need for deciding what the proper value of N might be. The number of observations will affect the result nevertheless, for as the number of judgments increases, the proportions will be more precisely determined and the value of r will be more exact.

TABLE I

COMPARISON OF PARAMETERS CALCULATED BY "CONSTANT PROCESS" AND STIMULUS REGRESSION METHOD

$S_{p.50}$ = Stimulus when $P = 50$ per cent; Slope (b) = Measure of Dispersion

Source of Data	*Constant Process*		*Stimulus Regression*			
	M	*SD*	$S_{p.50}$	Slope (b)	r	$\sigma_{\text{est.}}$
Bressler (from Woodworth) (Without weights)	98.28 grams	7.37	98.32	7.31	.9946	0.59
Bressler (With weights)	98.27 grams	7.22	98.28	7.18	.9980	0.40
Urban *S* 1 (from Brown & Thompson) (Without weights)	99.34 grams	5.30	99.36	5.32	.9988	0.39
Urban *S* 1 (With weights)	99.68 grams	5.20	99.37	5.03	.9800	1.11
Kellogg *S* 1 Table XVIII (Without weights)	3.259 step-units above lowest	1.52	3.249	1.46	.9787	0.411
Kellogg *S* 2 Table XVIII (Without weights)	2.502 step-units above lowest	4.46	2.526	4.25	.9767	0.429

Sampling errors in measurement will have a rather surprising effect on the values obtained by the proposed method. These will obviously be subject to the well-known regression effect. The lower the value of r, the more the predicted values will approach the mean of the stimulus series. Further, errors in measurement will, by attenuation, lower the value of r. Hence, error of measurement will reduce the predicted stimulus values, and a low prediction for *SD*, for example, would mean either a very good discrimination or very inaccurate measurement. This would be unfortunate, indeed, if it were not possible to distinguish between the two by looking at the standard error of estimate. In other words, comparison of experimental results would always need to be made with reference to the standard errors, which fortunately are available.

The conventionally derived values are not free from such an effect either. The normal equation for *SD* (in unweighted form) is

$$SD = \frac{N\sum x^2 - (\sum x)^2}{N\sum xy - \sum x \sum y} i$$

where i is the step interval, N=number of stimuli, x=stimulus value, $y=SD$ value. Since the value of x is presumed to be subject to no error, the numerator may be regarded as a fixed value for a particular experiment. The denominator is the familiar sum of the cross-products as found in the numerator of r, as far as the variables are concerned. Random errors of measurement would tend to depress the value of this sum, hence inflate the value of *SD*, and inflate it in an accelerated fashion. On the other hand, the reciprocal of *SD*, or the old-fashioned h, which has a normal equation with the fraction inverted, would be depressed by error, in direct proportion to the depression of the corrected sums of squares.

The regression effects for the conventional mean are more complex. Its equation is

$$M = \left(\frac{\sum x \sum xy - \sum x^2 \sum y}{N\sum xy - \sum x \sum y}\right) c + M_x.$$

The cross-products term in the numerator would somewhat offset the one in the denominator, but the net effect would seem to be an "outward regression."

To interpret a standard error, it is obviously desirable that it should represent a fairly normal distribution. When the stimulus regression method is used, it is assumed that the x-residuals have a normal distribution; it is consistent, therefore, to expect a normal distribution of the resulting parameters. The Müller-Urban process, on the other hand, in making a least-squares fit assumes that the y-residuals have a normal distribution, or, rather, if weights are used, assumes that underlying p-values are normally distributed. Now if normality exists in one of these three places p-, y (gamma or *SD*)- or x-values, it cannot be present in the other two. Consider the simplified case in which only two comparison stimuli have been used, yielding the data P_1 and P_2, each subject to a sampling error. The first step in computation would be to transform them to gamma or *SD* values by means of the normal probability tables. If we think of each of the points as being surrounded by a population, it is clear that the distribution of the population will not be the same after transformation as before. (This is an old observation [11].) Further: suppose we think of the distribution of gamma-*SD* values about a given point as represented in the third dimension, z, rising from the surface of the paper. There will be two of these distributions corresponding to the two data points. Fitting lines in this case would mean simply connecting pairs of points, and if we connect all possible pairs of points in each layer of the z-distribution, we shall generate a solid in the shape of a molding. If the y-values (gamma-*SD*) are normally distributed, the molding would display a normal curve when cut by a plane along one of the ordinates. If, on the other hand, the x-values (stimulus) are normally distributed, it would show a normal distribution when cut by a vertical plane on the x-axis. The cut cannot be normal on both axes except in a special case. If one is normal, the other varies according to the slope of the median line, from a rectangular distribution to one more sharply peaked than normal.

So if the p- or y- values are normally distributed, as assumed in the Müller-Urban process, it would be inconsistent to expect the derived parameters, which are x-values, to be normally distributed. As stimulus values seem a more likely place to expect a normal distribution, we have an additional reason for adopting the stimulus regression form of computation. The assumption of normality for p-values is recognized as dubious anyway, since, on the basis of the binomial expansion the sampling distribution of p's is not normal unless p is $\frac{1}{2}$ or N very large. It is, of course, a question to be decided by experimentation whether stimulus parameter values, as secured by the stimulus regression, are normally distributed or not.

It is no small advantage in an s-regression that the parameters are so easily computed. It would be possible to derive a pair of formulas like those of the Müller-Urban process but based on the s-regression. Considering the familiarity of the ordinary regression equation and the ease with which the necessary elements can be computed, there seems to be no advantage in such additional expressions. The computational procedure would be as follows, using the ordinary regression equation:

1. The p-values derived from an experiment would be converted into *SD* values by means of the normal probability table.
2. Taking each pair (stimulus value and *SD* value) as an entry in a correlation, the r would be computed. (The two means and σ's would be by-products of this calculation.)
3. The values of σ, σ_x, σ_y and M_x would be inserted into the x-regression equation, and the equation would be solved for the value of x when $(y-M_y)=1$. This will give the most probable stimulus for the percentage of judgments 1 *SD* above the mean of the fitted ogive.
4. The regression equation is then solved for x when $y=0$. This will produce the most probable value of x for the 50 per cent value of p.
5. The standard error of estimate is then computed by the usual formula. It is the same for both predictions. If, however, the *SD* is to be considered as an absolute value rather than as a point on a scale, it is the slope of the regression line, and its error is the error of the regression coefficient.

In this method, as in the classical Müller-Urban, weights could be used if the data are exact enough and the need for exactitude very great. With reasonably consistent data, however, experience seems to have shown they are hardly worth the trouble. Most experimenters will probably continue to dispense with them.

The stimulus regression method is not offered as the one and only method of solving the problem, but rather, as a sound practical expedient. Others may be developed which may suit particular purposes better, and it may be possible to develop a procedure which is more general in that it does not require knowledge of the form of the psychometric function or an approximately normal distribution of thresholds on the stimulus scale.

SUMMARY

Since experimenters need to know the sampling errors of the results they obtained, the traditional methods of calculating the psychophysical parameters from percentage data leave much to be desired. The several proposals previously made for calculating the standard errors of these parameters turn out to be incorrect, and it seems the parameters have such distributions that their sampling variation cannot be evaluated.

If, instead, one is content to state the problem as that of finding stimulus values most probably yielding designated proportions of judgments, the problem is quite easily solved. The regression of the transformed proportions on the stimulus scale is calculated, and the standard error is the standard error of estimate.

BIBLIOGRAPHY

1. Boring, E. G. The number of observations on which a limen may be based. *Amer. J. Psychol.*, 1916, **27**, 315–319.
2. Brown, W., & Thompson, G. *Essentials of mental measurement.* Cambridge, Eng.: Cambridge Univ. Press, 1921.
3. Bruner, J. S., & Postman, L. The reliability of constant errors in psychophysical measurements. *J. Psychol.*, 1946, **21**, 293–299.
4. Culler, E. Studies in psychometric theory. *Psychol. Monogr.*, 1926, **35**, 56–137.
5. Culler, E. Studies in psychometric theory. *J. exp. Psychol.*, 1927, **10**, 463–477.
6. Culler, E. The accuracy of the method of constant stimuli; a reply to Dr. Urban. *Amer. J. Psychol.*, 1927, **38**, 307–312.
7. Finney, D. J. The application of *probit analysis* to the results of mental tests. *Psychometrika*, 1944, **9**, 31–40.
8. Guilford, J. P. *Psychometric methods*. New York: McGraw-Hill, 1936.
9. Irwin, F. W. Psychophysical measurement methods. *Psychol. Bull.*, 1935, **32**, 140–171.
10. Kellogg, W. N. An experimental comparison of psychophysical methods. *Arch. Psychol., N. Y.*, 1929, No. 106.
11. Thompson, G. H. The accuracy of the $\phi(\gamma)$ process. *Brit. J. Psychol.*, 1914, **7**, 44–55.
12. Urban, F. M. Die psychophysischen Massmethoden als Grundlagen empirischer Messungen. *Arch. ges. Psychol.*, 1909, **15**, 261–355.
13. Urban, F. M. The accuracy of the method of constant stimuli. *Amer. J. Psychol.*, 1927, **38**, 252–256.
14. Woodworth, R. S. *Experimental psychology*. New York: Holt, 1938.

Received October 11, 1950.

BOOK REVIEWS

Gibson, James J. *The perception of the visual world.* Boston: Houghton Mifflin, 1950. Pp. xiv+242. $4.00.

In these stern days of harsh, quantitative, statisticized psychophysics it is with surprise that one turns up so fresh, so clear, so stimulating, and so unquantitative a phenomenology as is Gibson's scholarly description of what we see and how we see. The book's appearance becomes, in a sense, an event within a tradition: Goethe (brilliant, erudite, dogmatic, wrong); Purkinje (keen, usually right); Hering (argumentative, pedantic); and now Gibson creating the paradox of a phenomenology of vision which is both modern and American.

The account is indeed a phenomenology. There is not a table in this book, nor a graph, nor an equation, nor a quantitative functional relation, nor for that matter a number, except the page numbers, the numbers of the figures, and a few little numbers like *three*, which is the number of dimensions in Euclidean space. The author's critics will note all this and call him unscientific. Very well. Perhaps this is all *Vorwissenschaft* and not science. Who cares about words? The talk is good, penetrating discussion. The argument alters certain basic conceptual views about visual space perception. The discussion asks you—as phenomenology so often does—to become aware of the obvious, to shed many of the hard shells of classical pedantry and to accept instead the author's convincing demonstrations, so many of which are printed right there on the page for you to see.

The eighty-one figures (half-tones and line drawings) are magnificently good. No one can study them without seeing how insistently tridimensional is monocular vision, and perhaps the reader will go all the way with the author (and thus with R. S. Woodworth, too) to decide that the old secondary cues to depth are really the primary ones, and the classical primary cues but secondary.

It is natural and proper for a phenomenology to be descriptive, and Gibson describes, not merely *what* you see, as the Gestalt psychologists do, but also *how* you see. He is analytical. Description has to be analytical. Yet his terms of analysis are not points in a retinal or visual field, or even lines and areas, but relationships, orders. The visual field, he shows you, is ordinal, and it is these topological relationships which are the substance of visual perception. Perceptions in which the ordinal relations are invariant are identical, Gibson notes, using this means for generalizing the various facts of perceptual transposition and of behavioral response to a relationship.

Inevitably the study of perception forces one to face the dilemma of experience vs. knowledge. You look down the long, straight railroad tracks. They are parallel (knowledge) and they converge (experience). You can *see* the parallels converging. You can see both that the separation does not diminish with distance and also that it does. Faced by this contradiction, modern American functionalism chooses knowledge and lets experience go. If you count on the tracks' being just as far apart away off there as they are here, you will not get in trouble, that is to say, you will be right. Still, you can see them converging. What science takes account of that fact if psychology rejects it?

Of this dilemma Gibson takes a firm grip on both horns. He establishes two correlated systems: the *visual field*, the straight geometry of immediate visualization, and the *visual world*, which consists of the pattern of objects that make up perception. We shall do well to elaborate the specifications of these two systems: the field and the world.

The visual field is *seen*; the visual world is *known*. Thus the concept of the field is the modern descendant of the concept of *sensation*, whereas the concept of the world is descended from the concept of *perception*. In the field parallels may *converge*; in the world they remain always *equidistant*. The field tends to be *areal* and the world *tridimensional*, yet that distinction is not inevitable, for a surface may be an object in the visual world. It is better to say that the field is *pictorial*, representing (but not exactly copying) the *anatomical* pattern of retinal excitation, whereas the world is *objective*, constituted of *ordinal* relations. The visual field is *limited* in extent, is ever *changing* and *fluid*, and is consequently *non-Euclidean*, whereas the visual world is *unbounded*, *stable*, *upright*, *rigid*, and the dear old *Euclidean* space which is what we like to call real. The visual world is thus the *natural* world and the immediate substance of perception. The visual field, which the Wundtians used to think was immediate experience, Gibson believes can be apprehended by a kind of *introspection*, a sophisticated abstraction from objective reality to the system which resembles the retinal excitatory pattern, an abstraction in which artists and introspective psychologists are more adept than others.

When the Gestalt psychologists substituted objects (phenomena, meanings) for the older Wundtian sensations, they were saying that the objective visual world is the primary stuff of experience and that the sensations which introspection yields are the secondary sophisticated artifacts of a special kind of observation. Gibson follows the Gestaltists with respect to the objective visual world: objects are what human adults, children, and at least the higher vertebrates have in their perceptions ("size constancy" holds for chickens). So the one school inverts the other. For the Wundtians, sensory experience is primary and objects are generated from it. For the Gestaltists, objects are the stuff of experience and primary, with sensations becoming psychologists' constructs abstracted from the objects. Gibson's view of the relation of the field to the world is not clear to the reviewer. He seems to think of the field as primitive and the world as primary, of the field as genetically prior to the world and of the world as temporally prior to the field in observation. You cannot be quite sure whether the revelation of the field to introspection is an atavistic process or an act of sophistication.

The greater part of Gibson's book is the description of the way in which the visual world of objects depends upon the visual field, while differing in so many of its properties. The basic demonstration shows how the tridimensionality of the visual world is the consequence of certain ordinal relations in the visual field, relations of which one of the most important and most novel is the *gradient* of *texture*. Gibson constantly stresses gradients as conditions of objective perceptions because gradients show how the ordinal properties of an entire field provide the essential basis for the spatial orientation of perceived objects and for the organization of the objective world as a whole. Texture is large grain or macrostructure, tiles in a pavement, clumps of grass or soil in a field, stones in a gravel pit, anything that is constantly repeated in units of about the same real size. Just as the railroad tracks converge in the distance of the visual field, so the grain of the ground gets more finely structured in the distance, and the gradient of change in texture tells you, not only how fast the ground spreads away from you, but also how far below you it is.

Gibson's epistemology is correlational—like Mach's, or, more recently, B. F. Skinner's empty organism. What Gibson calls a "theory" is thus only a description of a correlation, a theory which tells *how* but skimps on *why*. That is, of course, all right for phenomenology, provided the enthusiastic describer does not forget that eventually science must go deeper into the means of correlation, must show in physiology *why* a gradient of texture produces a perceived depth, not merely that it does.

The social psychologists will be piqued to note how well Gibson gets along without using the concept of set, attitude, motive. He has, nevertheless, to come to the dynamics of perception in the end, as he does in his chapters on meaning and learning, and this reviewer thinks he would have fared better had he not resisted the inevitable so long.

There ought to be some accepted device by which anachronisms could be avoided in bibliographies. In this respect Gibson is no better than others. He gives the date of a translation but not the date when the thought was first published. When he has used a later edition he gives its date, which makes his page reference accurate and misleads the student as to when some thought was averred. Thus the careful student who wonders enough about history to examine dates in the bibliography will be misinformed about Helmholtz, Mach, LeConte, Berkeley (who "wrote" for Everyman in 1922!), and a few others.

It is too bad that the reproduction of the text of this book is inferior to its splendid figures. The verbal matter is the offset reproduction of typing. The spacing is poor, much poorer than is required by this method.

It is interesting to see phenomenology asserting a role in modern American psychology. To some extent Gibson rides the Zeitgeist. Retinal disparity and convergence he relegates to secondary positions, and for accommodation he has little use. If, however, the Zeitgeist is headed for the behavioristics of operationism and for a monistic unity of science, then he lags behind progress. He should have removed all doubt about the distinction between the visual field and the visual world by describing the operations by which each can be observed in chimpanzees. That would be, however, a good deal to ask of a phenomenology, which has a right to be just what it claims to be and not something else; and, moreover, how can we be sure that the Zeitgeist is set eventually to make behavioristics the indisputable scientific truth? All in all the reviewer finds this book a remarkably keen, clear, and wise description of just how it is that people see things. The chimpanzee phenomenology of vision, moreover, may need to wait until we have trained a chimpanzee to write it.

EDWIN G. BORING.

Harvard University.

HARRIS, R. E., MILLER, J. G., MUENCH, G. A., STONE, L. J., TEUBER, H. L., & ZUBIN, J. *Recent advances in diagnostic psychological testing.* Springfield, Ill. Charles C Thomas, 1950. Pp. x+120. $2.75.

This symposium of six contributions opens with an introductory essay by Miller on contemporary trends in psychodiagnosis and proceeds through contributions by Harris on psychodiagnostic testing in psychiatry and psychosomatic medicine, by Teuber on testing in neuropsychology, by Muench on the uses of diagnostic psychological methods in counseling and psychotherapy (including their validation), by Stone on recent developments in diagnostic testing of children, and a final chapter by Zubin on test construction and methodology. The book is even more brief than its 120 pages might indicate since 15 of them are taken up by bibliography, the pages themselves are small, and the type face is reasonably large. It would, perhaps, have been better presented as a monograph. The contributions are excellent, however, and they are well integrated, with a consequent absence of annoying overlap.

Miller's introduction is competent and timely; Muench's contribution deals interestingly with a not too often discussed facet of testing; and the Teuber and Zubin articles are able and stimulating performances in special technical fields. In this reviewer's opinion, special praise should go to the papers by Harris and Stone, where demonstrated clinical sensitivity goes hand in hand with theoretical competence and thorough appreciation of experimental methodology. They are a pleasing confirmation that the psychologist can function clinically without sacrificing his scientific heritage and that the blending of these two influences can be a promising and possible reality, not a mere ideal fiction kept alive by the stereotypes of our present professional training programs.

WILLIAM A. HUNT.

Northwestern University.

WORTIS, JOSEPH. *Soviet psychiatry.* Baltimore: Williams & Wilkins, 1950. Pp. xv+314. $5.00.

Outside the borders of the U.S.S.R. knowledge of Soviet psychiatry and psychiatric practice has been excessively meager. Dr. Wortis in his book *Soviet Psychiatry* has tried to reduce this lack by presenting, against an historical and ideological background, a report and modest analysis of the contemporary state of the Soviet psychiatric scene. The book covers a considerable range of material, but fails to produce the kind of account we should have liked.

There are a number of basic reasons for this failure. They can, for the most part, be summarized in the author's uncritical relationship to the language and content of his material which, even if adequate in coverage (and it is not), would be insufficient basis for the task the author set for himself. The realities of the U.S.S.R. are not to be plucked directly from a literature which is commanded to eschew "objectivism" and to exhibit the requisite "partisanship" demanded by the Party. Because the Soviets speak of "free, creative discussions" does not mean that there *are* free, creative discussions; nor must we necessarily acknowledge the Pavlovian basis of contemporary psychiatry in the Soviet Union, merely because the "glories of Pavlovian theory" are dutifully intoned in the prefaces of books and the columns of the press.

Yet the author of *Soviet Psychiatry* does precisely this. The declarative allegiance to Pavlovian theory in much of Soviet psychiatry is not perceived as such, and the model, nonexistent "Soviet psychologist" and "Soviet psychiatrist" are made the uniform spokesmen of the living flesh and blood. Thus, from the author's account one would not gather that Pavlovian theory was ever termed in the U.S.S.R. a "verbal husk" and a "huge mechanistic peril" to Soviet psychiatry; nor would one have suspected from his account that there could be ground for saying, as Ivanov-Smolenskiĭ did in a meeting called in 1950 to introduce *in fact* "Pavlovianism" into Soviet psychiatry, that "until recently all [real] attempts at application of Pavlovian theory to the problems of psychiatry have been bayoneted invariably."

This book is further marred by unnecessary excursions into Soviet apologetics. "Forced labor camps" are said to demonstrate the Soviets' "deep confidence in the wholesome restorative functions of work." And the disgraceful Lysenko affair, whose gratuitous introduction into the book serves no useful purpose, is unrealistically excused away.

One could refer to other serious defects, such as the failure on the author's part to understand that references, distributed over a period of years, cannot be amalgamated at will to yield a composite contemporary picture. However, with the proper reservations in mind, the circumspect reader will find much of interest and value in this book on a subject little explored outside the U.S.S.R. The book is a pioneer attempt in a field where the pitfalls are many. For this reason alone the author is to be commended for venturing on so difficult an assignment. However, the book which can be said to have presented Soviet psychiatry and psychiatric practice adequately still remains to be written.

IVAN D. LONDON.

Russian Research Center, Harvard University.

VON FRISCH, KARL. *Bees: Their vision, chemical senses, and language.* Ithaca: Cornell Univ. Press, 1950. Pp. xiii+119. $3.00.

As indicated in its title, this thin volume (119 pages) describes the visual and chemical senses of the honeybee. The evidence, much of it gathered by von Frisch over the past thirty-five years, is presented clearly and interestingly in nontechnical language.

The bee's color sense varies from that of human beings. Ultraviolet radiations are visible to this insect, but it is not sensitive to red. Form perception depends upon the "brokenness" of a pattern, and this may be related to the fact that bees use vision primarily while in flight. Von Frisch suggests that "broken" patterns produce a flickering sensation. The olfactory sense of the honeybee is closely comparable to that of a man "whose sense of smell is very well developed." Bees differentiate between sweet, salty, sour, and bitter tastes but are much less sensitive to bitter substances than are human beings.

The most interesting part of the book is that which describes what the author prefers to call "the language of the bees." As a result of many years of careful, ingenious investigation von Frisch is able to present evidence of which he says, "No competent scientist *ought* to believe these things on first hearing." Those who were fortunate enough to hear Dr. von Frisch lecture during his visit to this country in 1949 may have shared with this reviewer the conversion from skepticism to enthusiastic acceptance of von Frisch's published findings.

The gist of his reports is as follows. When a honeybee finds a new source of nectar and returns laden to the hive, she behaves in such a way that her hive mates leave the hive and fly to the location of the food source. The bees are influenced by chemical stimuli associated with the "discoverer" and also by the "dance" which she performs on the vertical surface of the comb. The pattern and tempo of her dance vary according to the location of the food supply and also according to the distance that must be traversed to reach it.

Von Frisch has found that the pattern of the dance is controlled by the plane of polarization of diffused sunlight. The degree of polarization in any portion of the sky at a given time is a function of the sun's position and the structure of the bee's compound eye, particularly the structure of the single ommatidium which appears adapted to function as an analyzer of polarized light. The direction in which certain portions of the dance are oriented varies according to the direction of the flight path to the food. The speed with which parts of the dance are performed is inversely related to the distance flown.

By reacting appropriately to the cues mentioned above, bees that have never visited a particular food source are enabled to proceed directly from the hive to the new supply of nourishment. In what appear to be well-controlled experiments bees were observed to fly as far as 6 kilometers (3.7 miles) straight to a new feeding place from which a hive mate had recently returned loaded with nectar.

This book can be strongly recommended to readers who are scientifically concerned with comparative psychology. But it will also prove entertaining and instructive to laymen who are interested in natural history. The literary style is simple and charming. The subject matter stands as a model of patient, careful investigation of complex patterns of animal behavior.

F. A. BEACH.

Yale University.

BROWN, WARNER, AND GILHOUSEN, H. C. *College psychology.* New York: Prentice-Hall, 1950. Pp. xviii+507. $3.75.

The character of this text is set in two remarks by the authors in the early chapters of the book. "Psychology," they write, "is the study of behavior and experience." A bit later they observe, "So, instead of presenting theories, we continue in the next three chapters, as we began, maintaining an externally descriptive attitude." This last applies to the treatment of learning in particular, but could just as well apply to the whole work. Reading these, one would expect a book which is eclectic, nontheoretical, and rather traditional, and so it is.

The bifurcate definition, of course, enables the authors to stand with one foot on each of the familiar foundations: reflexology and the analysis of consciousness. In practice, indeed, the body of doctrine becomes quadripedal, for in the last fourth of the book, the authors write a psychology in terms of adjustment, then a psychology in terms of abilities. By and large, the authors seem to find the relations between the four different limbs too complex and uncertain to discuss in such a book.

The "descriptive attitude" the authors favor means, no doubt, predilection for facts and a postponement of theory. One can heartily praise an eclecticism of facts, but a jumbling of theories is not so clearly admirable. Granted that it could be done at all, the exposition of facts picked clean of theory would be a most difficult operation, and the result not likely to be palatable to first-year students. So it must be expected that authors contemning theory will, nevertheless, present their facts embedded in theories—diverse and sometimes very aged theories taken unwittingly from the general heritage.

For example, when the authors write, "what we ourselves contribute to the operation of thinking . . . consists of the ideas, memories, concepts, and images that 'pass through our minds,' " they are hardly reporting a simple untrammeled observation but, rather, repeating a very old theory. So, in "attending" (which is the "selective process" and the "awareness of concentration" "prior to perceiving") "the self," "adjustment," "motivation," etc., there is evident a matrix now of one psychological theory, now of another. As one goes through the book these are found intermingled. Simple learning is a matter of conditioned response. In trial-and-error learning there is activity ("trying to accomplish something"); in problem solving there are in addition "ideas," and thinking is wholly a matter of "ideas." If psychological theories are a vice, as Brown and Gilhousen maintain, "because they obscure the important distinction between facts that are actually known and a fanciful veneer of inventions covering the enormous area of the unknown," they have, indeed, taken a strange path to virtue by intermingling more than the usual number of different theories.

In the introduction, the authors announce that they do not expect the student to weigh or interpret experimental evidence but will cite experiments, when necessary, to make their statements concrete, choosing studies as old as possible so as to impress upon the student "the truth that psychology has historical foundations." It is clear that they themselves have great respect for psychology with historical foundations, and they would probably be well pleased to note the resemblances of their book to those written thirty years ago.

R. C. DAVIS.

Indiana University.

WIENER, NORBERT. *The human use of human beings.* New York: Houghton Mifflin, 1950. Pp. 241. $3.00.

The author of *Cybernetics* has written a related book for the layman to emphasize the "social consequences of his point of view." The issues treated can scarcely be traced to a single point of view, but they are an interesting and vigorous lot. There are discussions of the direction and nature of human progress, the distinction between rigid instinctive behavior and modifiability, man's exclusive claim to speech, the nature of the individual, the paradoxes of the law of patents, the evils of scientific secrecy, the education of the scientist, the second industrial revolution, the Catholic Church as a totalitarian institution, and so on. Although "thinking machines" are frequently mentioned, and many parallels are drawn between them and living organisms, few if any of the author's conclusions are derived from cybernetics itself in any rigorous way. It is a book of strong, often violent, opinion, ranging over many of the fields which an intellectual with compelling political and moral persuasions might be expected to touch upon in the course of an active life. Its manner swings from the stiffly academic to the witty, from the dogmatic to the humble, from the exaggerations of science fiction to the caution of the trained scientist.

Psychologists will not find much that contributes to their science. The author has brushed lightly against physiology, particularly the works of Sherrington and Cannon, but whenever his comparison of thinking machines and living organisms becomes in any sense concrete, the assumed mode of operation of the nervous system has a shopworn look. His acquaintance with behavioral processes is even less extensive. For example, although proprioceptive feed-back from behavior is one of the cardinal issues in *Cybernetics*, there is only the most superficial account of what such a feed-back is like, and no discussion whatsoever of experimental work in the study of learning in which the effect of feed-back upon behavior is analyzed.

In spite of this, the book has a certain value in reminding us that scientific achievements are seldom without an impact upon civilization. The special impact with which Professor Wiener is here occupied [illegible] certainly not received the attention it deserves.

B. F. SKINNER.

Harvard University.

Vol. 48 No. 5 September 1951

Psychological Bulletin

THE FREUDIAN THEORY OF DAYDREAMS: AN OPERATIONAL ANALYSIS[1]

WILLIAM SEEMAN

University of Pennsylvania

INTRODUCTION

The formulation of psychodynamics of daydreaming by Sigmund Freud (**9**, **10**, **11**, **13**) enjoys a distinctive and interesting status in the domain of behavior science which is demarcated as "clinical"; for, following Freud, it is almost universally accepted in clinical circles and in the clinical literature that daydreaming serves a gratification function—i.e., that it is wish-fulfilling in character. Indeed, so strong is this conviction with respect to the psychodynamic function of the daydream that in at least one instance (**41**) it is incorporated into the *definition* of daydreaming, which is said to be "the imaginary representation of satisfactions that are not attained in real experience" (**41**, p. 186). Yet, notwithstanding the strength of this conviction and the widespread acceptance of the theory, it does not appear that it has ever been empirically demonstrated. Indeed, there is apparently no formal or operational analysis in the literature which indicates *how* the theory may be laid open to empirical investigation, and consequently to verification—to confirmation or disconfirmation. Such an analysis is nevertheless an indispensable prerequisite to empirical investigation. As it stands, the theoretical formulation is inadequate in operational specifi-

---

[1] This paper is adapted from a thesis presented to the Psychology Department of the University of Minnesota in partial fulfilment of the Ph.D. degree. I am indebted to Dr. Laurence F. Shaffer, of Teachers College, Columbia University, for time given and for his generosity in permitting me to see some of his unpublished daydream data; to Dr. Wallace A. Russell, whose incisive comments and criticism aided materially in clarifying certain formal aspects of the analysis; to my brothers, Dr. Julius Seeman, of the University of Chicago, and Dr. Melvin Seeman, of the Ohio State University; to Dr. Kenneth MacCorquodale, who in the course of three years of contact has contributed in no small measure to my thinking on this problem; and, finally, to Dr. Paul E. Meehl, whose encouragement and support were of inestimable value in the execution of the thesis investigation.

Psy. Research

TABLE I

SUMMARY OF LITERATURE RELEVANT TO THE FREUDIAN THEORY OF DAYDREAMS

*Author**	*Subject*	*Methodology*	*Quantification*	*Conclusions*
Bose (2)	Use of daydreams for therapy in psychoanalysis	Report and interpretation of use of fantasy with patients	None	Asks patient to "indulge in such daydreams as would give imaginary satisfaction to those [i.e., repressed] wishes" (2, p. 34)
Conklin (3)	Foster child fantasy	Questionnaire	Percentages	Foster child fantasy a "common" experience; reported by 24% of males, 31% of females
Conklin (4)	Relation between daydreaming and test adjustment	Personality inventory (Thurstone)	Percentages†	Relation between "neurotic" scores and "frequent" daydreaming
Dexter (5)	Imagination	Projective and structured personality tests; exciting events; sentence completions	Correlation	Correlations low positive most of them not significant; highest r (Bernreuter with self-rating) .527
Eidelberg (6)	Masturbation fantasy	Psychoanalytic interpretation of fantasy	None	Progressive changes in fantasy provide barometer for measuring therapeutic progress; may shorten therapy if understood
Ferenczi (7)	Gulliver fantasies	Psychoanalytic interpretations of *Gulliver's Travels*	None	"Compensatory wish-fulfilment" (7, p. 286) where fantasy involves size reduction
Fülop (17)	Relation of fantasy to body structure	Analysis of literature and clinical data	None	Fantasy life helpful or harmful, depending on morphology
Green (18)	Daydreams of child and adult clinical cases	Intuitive clinical analysis of case material	None	Accepts Freud's formulation but uses McDougall's concepts (e.g., daydreams gratify "gregarious instinct," and others (18, p. 19)
Hart (19)	Variety of fantasy	Interpretation of daydream material	None	Daydreams viewed as "building of pleasant mental pictures in which the complexes attain an imaginary fulfilment" (19, p. 155)
Hesnard (20)	Erotic daydreams	Psychoanalytic interpretation of sexual fantasy	None	Function of daydream: "Procurer . . . des l'excitations érotiques . . . plus ou moins intentionelles," (20, p. 524)
Hurlock and Burnstein (24)	Imaginary playmate	Questionnaire	Percentages	Imaginary playmate reported by 31% of females, 23% of males; 80% of these had strong positive affect; no instances of strong negative affect
Jaehner (25)	Imaginary companion	Interpretation of child fantasy	None	Inferiority feelings generate compensatory fantasy
Kamiat (26)	Cosmic fantasy	Speculative analysis of cosmic fantasy	None	Cosmic fantasy compensates for feelings of insecurity and inadequacy
Lehrman (29)	Variety of daydreams	Analysis of clinically observed behavior	None	Daydreams are wish-fulfilling; foster child fantasy "a compromise between incestuous wishes and the neurotic flight from it" (29, p. 343)
Lévy-Valensi (30)	"Bovarysme‡ et constitutions mentales"	Analysis of mental activity analogous to that of Madame Bovary	None	"Primitive" bovarysm not wish-fulfilling; "secondary" bovarysm "s'offre comme une *compensation*, une fuite dans la fiction pour échapper au réel" (30, p. 295)
Rosenzweig (38)	Methods of studying fantasy	Didactic discussion of difficulties and methods	None	"Infantility" of daydreaming stressed; child's method of securing "immediate satisfaction despite adverse circumstances" (38, p. 42)

TABLE I (Continued)

*Author**	*Subject*	*Methodology*	*Quantification*	*Conclusions*
Ruenaufer (40)	Variety of daydreams	Questionnaire; daydream descriptions of 65 Indian students	Percentages	Grandeur, personal exploits, and political freedom daydreams reported; some sex differences noted
Shaffer (41)	Variety of daydreams	Questionnaire	Percentages, means, and sigmas	Daydreams a mechanism of defense and source of satisfaction for the normal individual as well as the maladjusted
Smith (43)	Variety of daydreams	Description of daydream types	None	"Mental device for encompassing all desires"; daydreams relevant to needs; vary with sex and socioeconomic status
Varendonck (48)	Fantasy activity of author	Introspective analysis of genesis, content, termination of own thought chains	None	Daydreams "practically a search for pleasurable representations" (48, p. 15); differs from "logical ideation, which corresponds to reality" (48, p. 13); "directed by one or several wishes" (48, p. 276)
Van Waters (47)	Imaginary companion of a little girl	Analysis of fantasy and clinical material	None	Fantasy activity regarded as providing substitute gratification
Vostrovsky (49)	Imaginary companions	Questionnaire	Percentages	Imaginary companion may appear as late as adolescence
Woodworth (53)	General nature of fantasy	Analysis of daydream types appearing in literature	None	Daydream function: "getting for the moment the satisfaction of some desire" (53, p. 494)

* Listed in alphabetical order.

† Although Conklin reports only percentages and does not submit his data to tests of significance, he provides enough information to permit this. The computed χ^2, disregarding the contribution of one cell which fails to meet the criterion of theoretical $N=5$, is 59.19. This is significant beyond the 1 per cent level.

‡ Defined as "le pouvoir à l'homme de se concevoir autre qu'il n'est" (30, p. 289).

cations which would make it possible to derive such hypotheses as might lend themselves to quantitative confirmation within specified probability levels.

Some idea of the extensive acceptance of Freud's theory of daydreaming may be indicated by Table I, which presents in summary form a review of a good deal of the literature relevant to the theory. The column marked "Conclusions" indicates the frequency with which the gratification function assigned to the daydream by Freud is accepted by writers on the subject. One might get the impression that many clinicians suppose the theory to be rather solidly grounded in empirical evidence. Yet Conklin's lament that "of systematic studies we have not so many" (4, p. 217) may be echoed today, fifteen years later. As the table suggests, there have been few attempts at quantification: only one of the studies (5) employs correlation techniques; five (3, 4, 24, 40, 41) report percentages, and of these only Conklin's (4) and Shaffer's (41) data permit the calculation of tests of significance.

A critical evaluation of the comparative merits of these writings would require a careful and explicit distinction between the fruitfulness of *content* and methodological soundness and rigor. It should be clear that a formulation may be rich on the content side (i.e., potentially contain many hypotheses), but at the same time may suffer from severe methodological weaknesses. This would appear to be the case with Freud's formulations, which provide insightful hypotheses of a remarkably penetrating caliber for empirical investigation, and which have, as Mowrer and Ullman (**35**) have noted, considerable pragmatic merit. With respect to most of the other psychoanalytic papers cited (**2**, **6**, **7**, **20**, **26**), however, it would appear that they contribute little that is new, and that they are notably deficient in recognition of the formal problems involved.

Conklin's earlier paper (**3**) provides objective support for the psychoanalytic contention that the foster child fantasy is a common one; but in view of his earlier citation of Karl Abraham to the effect that the fantasy is universal, he appears to strain his own evidence considerably in stating that "Comparison with the psychoanalytic presentation . . . results in both support and amplification of the generalization from psychoanalysis" (**3**, p. 20). His college study (**4**) is a significant contribution to the problem in that it demonstrates statistically highly reliable covariation of personality inventory scores with frequency of reported daydreams, the covariation being in the direction which the theory would appear to require. Shaffer's study (**41**) likewise provides evidence which supports Freud's theory. The data confirm Freud's assertion as to the normality of daydreaming and justify Shaffer's statement that "daydreaming is an exceedingly common and therefore, in the statistical sense, a normal form of behavior" (**41**, p. 195). They also have deeper theoretical significance in that they demonstrate quantitative differences which are consistent with the formulation.

Freud laid out the general psychodynamic nature of daydreaming in several of his writings. In 1900 he stated:

> A more thorough examination of the character of these day-phantasies shows with what good reason the same name has been given to these formulations as to the products of nocturnal thought—*dreams*. They have essential features in common with noctural dreams; indeed, the investigation of daydreams might really have afforded the shortest and best approach to the understanding of nocturnal dreams. Like dreams, they are wish-fulfilments . . . (**13**, p. 457).

Later, in 1908, in a paper on hysterical fantasies, Freud wrote:

> These phantasies are wish-fulfilments, products of frustration and desire; they are justly called daydreams, for they give us a key to the understanding of night dreams, the nucleus of which is nothing else than those daytime phantasies, but complicated and distorted (**11**, p. 52).

And in the same year, writing on the poet and daydreaming:

> Let us try to learn some of the characteristics of daydreaming. We can begin by saying that happy people never make phantasies, only unsatisfied ones. Unsatisfied wishes are the driving power behind phantasies; every separate phantasy contains the fulfilment (**15**, p. 176).

Again, in his *General Introduction*, Freud writes, " . . . daydreaming also is a mode of activity closely linked up with gratification, which is, in fact, the only reason why people practice it" (**10**, p. 117); and elsewhere in the same volume, "Now daydreams are literally wish-fulfilments . . . " (**10**, p. 117).

Although, as White (**51**) points out, some rather formidable problems compelled Freud in 1933 to state that dreams were *attempted* wish-fulfillments which, under certain circumstances, could achieve their ends only incompletely, there is some reason to doubt that he ever really abandoned the original form of his theory. For later, in 1936, we find him again writing:

> . . . but the isolated thought is found to be an impulse in the form of a wish, often of a very repellent kind, which is foreign to the waking life of the dreamer and is consequently disavowed by him with surprise and indignation. This impulse is the actual constructor of the dream . . . (**8**, p. 80).

It is not here contended that this view of the psychodynamics of the daydream was wholly original with Freud, nor was it his exclusive property. In fact, Freud himself (**11**) credits Havelock Ellis with it, and cites Breuer, Janet, and Pick as well. We have seen, too, how widely accepted it is in the clinical literature and in clinical circles, irrespective of other theoretical commitments. It *is* true, however, that the formulation is most systematically made by Freud, who, in fact, goes on to discuss its similarity to nocturnal dreaming and to the formation of neurotic symptoms. It becomes an integral part of his theoretical structure. An analogous situation may be found with respect to Postulate 9 in Hull's (**23**) system, a postulate on conditioned inhibition which is taken almost intact from Pavlov and incorporated as a part of Hull's theory.[2] In fact, a large part of Hull's formulations have an honored history in psychology; and this is only what one would expect in the biography of a science. This obtains no less in physics, where, prior to Einstein's formulations of relativity theory, other distinguished efforts were made to solve some of the paradoxical results of experimentation.

LEVELS OF THEORY

A scientific theoretical structure may be viewed as a language system; and, if the term "primitive" is used in the same sense in which it is used by Whitehead and Russell in *Principia Mathematica* (52), a relation of "more primitive" may exist between two language systems. The term "levels" is used here to convey the notion of this kind of relationship. In psychology the distinction between levels of analysis has been made in terms of molar versus molecular levels (23, 44, 45). The same set of behavioral events, as Spence (44) has pointed out, may be described in languages which are at different levels of analysis. In behavior theory this has apparently led, at times, to the notion that the two descriptions are at variance. About this, Spence writes:

---

[2] In his most recent statement (**21**), this is Postulate X.

> Such different descriptions, however, do not necessarily represent fundamental disagreements. If the two systems of concepts should each be successful in leading to the discovery and formulation of laws, it should also be possible to discover co-ordinating definitions which will reveal the interrelations of the two systems. Or, as Hull suggests, the postulates . . . at a more molar level may ultimately appear as theorems in a more molecular description (44, p. 71).

A further illustration of the concept of levels of theory is provided by Hull (22), and here we note that the distinction in levels is not made along molar-molecular lines. After elaborating a miniature system of adaptive behavior, consisting of eighteen definitions and six postulates, Hull develops a number of theorems. In Theorem XII he states: "*Organisms capable of acquiring anticipatory goal reactions will strive to bring about situations which are reinforcing.*" Such "striving" to bring about a goal state of affairs will be recognized at once as fitting the concept of "motivation," and Hull therefore states in a footnote: "An additional element of interest in this theorem is the fact that the fundamental phenomenon of motivation seems to have been derived from the ordinary principle of association . . ." (22, p. 14). Thus in the language system of learning theory the concept of motivation is regarded as derivable "from associationist principles,"[3] in contrast to its status at the "clinical" level, where motivation is taken as a "primitive" concept. Such distinction between the "learning" and "clinical" languages is involved in discussions by Mowrer (32), Miller (31), Mowrer and Lamoreau (34), Mowrer and Ullman (35), and Mowrer and Kluckhohn (33). These writers also attempt to discover relationships between the two language systems.

The attention which clinical psychologists have given to serious theoretical formulation has been, until recently, only moderate in degree. And it is perhaps this, as much as anything else, which accounts for the fact that clinical psychology has developed almost a separate biography as a behavior science, and that there has been minimal communication between systematic (academic) psychology and clinical psychology. Doubtless the intensely pragmatic demands of clinical psychology as an "applied" domain also account considerably for this failure of communication. Recently, however, clinical psychologists have been demonstrating interest in systematic formulations, and the possibility is even envisioned that clinical psychology might become "in considerable measure the content of systematic psychology" (**39**, p. 5). To anyone who regards it as desirable that psychology eventually develop the kind of formal and systematic structure which is perhaps a defining property of scientific maturity, this turning of attention on the part of clinical psychologists to problems of theory and of theory building at the "clinical" level must be regarded as a salutary development. And this is true not only because a body of coherent theoretical formulations has heuristic value in sharpening problems and indicating avenues and areas of investigation, but also because it compels a consideration of how investigation and theory in the clinical area can most fruitfully be incorporated into a unified discipline of behavioral science.

---

[3] It is not intended here to convey the notion that this has, in fact, been successfully effected. On this, see Koch (27).

Already there are a number of formal structures at the clinical level: a provocative and intriguing attempt at theory building at this level has been made by Murray *et al.* (**36**). In their well-known *Explorations in Personality* they have elaborated a system of theoretical constructs including need, press, thema, regnancy, and others. The system is frankly centralistic in character, has a certain amount of imaginative appeal, and constituted at the time of its publication a daring departure from the strict criteria of a narrowly conceived behaviorism. It remains, however, quite incomplete. Another attempt at theory building at the clinical level is that of Rogers, who has called his mimeographed form a "tentative draft" (**37**). The "theory" consists of a series of seventeen propositions, each elaborated at some length and dealing with: (*a*) the reaction of the organism as "an organized whole"; (*b*) the nature of his perceptual behavior; (*c*) the concept of the self; (*d*) the nature of psychological maladjustment, which is held to be a consequence of "a discrepancy between the organic perceptions and the self-concept" (**37**, p. 13); and (*e*) the effect of threat on the self-concept. Some attempt at systematic formulation is also to be found in Lecky's (**28**) little book, where the concept of self-consistency is central to the "theory."

The only formulation at the clinical level which, in my opinion, may be properly designated as a theory of behavior is that of psychoanalysis. This statement is not intended to convey any notion with respect to the "truth" value of psychoanalytic theory. It has reference to the comprehensiveness of the theoretical structure and to the relations which exist among the propositions. It is recognized that a theory may be most comprehensive and at the same time "false" in the sense that (*a*) it asserts propositions which are contradicted by empirical and experimental evidence, and (*b*) it asserts propositions which are mutually contradictory. Unfortunately, the most cogent statement of the system in rigorous terms has not yet been made, and some of its adherents and practitioners admit that much psychoanalytic writing is fluid, ambiguous, and unintegrated. To repeat, therefore: there is no intent here to assert that psychoanalysis is the only "true" system of psychodynamics, whatever that might mean. The reference is merely to the fact that Freud appears to have been the only theorist who has laid out with some comprehensiveness and care an explanation on the clinical level as to the nature of human behavior.

OPERATIONAL ANALYSIS

An integral proposition in this framework of psychoanalytic theory is that which assigns to the daydream the general psychodynamic function of wish-fulfillment. It will be my purpose in the remainder of this paper to present a formal analysis of the problems presented in an empirical verification of this theory and to suggest a "methodological model" for an investigation. It appears to be the primary objective of such an analysis to indicate the defining operations which are indispensable to casting the theory into a form susceptible of quantitative empirical investigation.

It seems appropriate, in connection with this problem, to consider first the significance of the *wish language*.[4] If one sticks close to Freud's intent, it would appear that the definition of a wish should be in terms of goal objects and/or goal states of affairs. While Freud's "Wünsche" *could* probably be developed in terms of some *drive* language, and while such a development could be justified on pragmatic grounds if it turned out to be more fruitful, that would still constitute some departure from what Freud appears to have in mind in his use of the wish language.[5]

---

[4] The choice of the word "wish" in the translation has more than trivial consequences, and it is therefore important to note that it is an accurate rendering. In the original German we find "Diese Phantasien sind Wunschbefriedigungen . . . " (12, p. 192); and "Unbefriedigte Wünsche sind die Triebkrafte der Phantasien . . . " (9, p. 216). The significance of this lies in the implications of the wish language which will be developed.

[5] The use of a drive terminology would present other problems as well. For one thing, as Skinner (42) points out, there is no *one* conceptual drive formulation in psychology which is accepted by all psychologists. In some instances it is regarded as a *stimulus*

Such a departure would appear to be unnecessary, for Tolman (**45**) has developed a concept, the concept of "demand," which may be defined in terms of goal objects and/or goal states of affairs. In some detail, he presents experimental operations "indicating the reality and objective definition of the rat's 'demand for' specific types of goal objects" (**45**, p. 37). As developed in his book, however, the concept is somewhat limited for the purpose of the present task. It needs to be expanded to include more complex states of affairs—e.g., "demand for vocational success," "demand for personal attractiveness," demands for sex and food objects. Such expansion presents no major formal problems. The defining operation for any demand (wish) can be explicitly described in terms of verbalizations of a specified kind; e.g., a "demand for success" might be defined in terms of verbalizations with respect to the acquisition of certain kinds of objects. But there is no logical reason to assume that the defining operations must be *limited* to verbal behavior. We may at the present time leave it an open question whether a concept like "unconscious demand" ("unconscious wish") will be necessary, or whether a defining operation can be found for such a concept.

Given specific defining operations for specific types of wishes (demands)—such as demand for a sex object, for "vocational success," for "physical attractiveness," etc.—the Freudian theory that daydreams are wish-fulfillments may be restated as follows: *The emission of a daydream is functionally related to a specific type of demand (wish), the relation being such that whenever an instance of such and such a daydream is observed, it is required by the theory that an instance of a specified corresponding demand (wish) must be identified by a suitable objective operation.* It seems clear that, so stated, the theory really requires the occurrence of identifiable, lawful patterns of *demand-daydream* covariation. What is crucially important here is the understanding of the *contingent notion of frequency*, which lies buried in this analysis of the *meaning* of the concept of wish-fulfillment. A concept of "frequency" is involved in the *definition* in the sense that there is implied the notion that whenever an instance of a daydream is counted an instance of a relevant demand must be identified. That this is *not* an assumption apart from the *definition* of wish-fulfillment, but a contingent condition which it imposes, is the essential point at issue here. The significance

---

(**41**); this would pose great difficulties. In others (**21**), the stimulus properties are predominant, though not exclusive. For most psychologists its physiological correlates appear to be crucial. Finally, the use of a drive terminology might conceivably pose formal problems in an attempt to distinguish between primary and secondary drives.

of this point lies in the fact that specific deductions from the theory will be essentially *predictions of where these frequencies may be expected to lie in consequence of theoretical requirements.*

It is necessary also to introduce explicitly the concept of "stronger than" as a relation between demands. This can be done by definition, and in any one of several alternative ways, each of which would be adequate to the purposes of the present task. Tolman's own definition, as introduced in the following passage, is quite acceptable:

> Turning now to the results, it appears that the groups may be arranged in the order of goodness of their performances. It appears, in short, . . . certain goal objects or situations produce better total maze performance than do others. And this introduces us to the conception that certain goal objects are . . . demanded more than others . . . . The strength of the demand for the type of goal object provided is thus one of the immediate immanent aspects inherent in, and defining itself through, maze performances (**45**, p. 41).

This defining operation has certain similarities to Warden's methodology (**50**), in which the drive-measuring operation is likewise essentially a rank-ordering with respect to performances. However, there is available a convenient alternative procedure, more nearly isomorphic with the defining operations for "denser than" cited by Bergmann and Spence (**1**, p. 8) in which liquid Y is said by definition to be denser than liquid X if X floats on Y. Such an operation has already been performed by Tsai (**46**) in comparing the relative strength of sex and hunger motives. The analogous operation with respect to the present problem would be as follows: where it has been empirically demonstrated that a choice is made at the 1 per cent level of confidence between two objects or states of affairs, the relation of "stronger than" will be said by definition to hold for the specified demands.[6]

The formal aspects of an investigation of this character, which aims at empirical verification of a theory, ordinarily takes the form of an inductive leap from $(P \supset Q) \cdot Q$ to (P) where P is the theoretical formulation (in this instance "daydreams are wish-fulfillments") and Q a class of observable behavioral events (e.g., responses).[7] However, it frequently happens that a proposition or predicate (Q) is entailed not by a single proposition but by several, and the symbolic statement in the case of the problem under consideration would take the form

---

[6] In selecting such a procedure as a defining operation, however, it must not be forgotten that there is no *logical* reason to assume a transitivity relationship. That is, empirically, it is at least possible that A can be in greater demand than B and B in greater demand than C without assurance that A will be in greater demand than C. Should this turn out to be the case, the defining operation will have proved inadequate, and it would then be necessary to introduce the concept in a manner more nearly isomorphic with the Tolman and the Warden procedures.

[7] This is, of course, an extremely elementary application of symbolic logic. The symbol "$\supset$" is the symbol of *implication*, and the expression $P \supset Q$ is read "if P, then Q," or "P implies Q." The symbol "$\cdot$" indicates a "binding" operation, and is usually

$$[(P \cdot P'_{1 \dots k}) \supset Q_{1 \dots k}] \cdot [Q]$$

where P represents the major proposition (i.e., the theory) under investigation and $P_1{}'_{\dots k}$ represents in each instance either a known empirical fact about the behavior of organisms (e.g., a known fact about adolescent interests and preferences) or an assumption.[8] The several propositions, then, P and P', entail a consequence, a predicate, Q; and it is this logically derived consequence which will be, in each case, the predicate which constitutes the hypothesis (H) which the empirical data would be required to confirm or unconfirm. In instances where the confirmation conditions show Q to be in fact the case, there is nothing embarrassing about this; but where the confirmation conditions fail to do so, and $\sim Q$ is the case, then there are two sources of possible error, and it would be difficult to determine whether the source is P or P'. The argument here would be, however, that if in a large number of instances (i.e., test conditions) involving the same P but different P''s there is overwhelming confirmation of Q, then in those isolated instances (should such occur) where the confirmation conditions indicate that $\sim Q$ is the case, the indication would be for a re-examination of P' before P.

In conclusion I should like to illustrate the procedure under this "methodological model." This may be done by examining some of the data presented by Ruenaufer (**40**), already cited in Table I. He informs us that there is a measurably different intensity between male and female college students in India with respect to the daydream of independence. It could undoubtedly be established by objective operational procedures that the demand (wish) for independence is more "masculine" in the sense that it is so rated by *both* males and females in India.

---

read as "and." Whether because of the advantages of clarity and precision of deduction, or because of other advantages, symbolic logicians have apparently extended the power of analysis beyond that which has been characteristic of the classical logic. Those interested in a more complete description of the nature of symbolic logic are referred to Whitehead and Russell (**52**).

[8] An alternative way of stating this would be that P' represents a class of propositions which constitute statements about known empirical facts or about assumptions. Since, in the event of the occurrence of confirmation conditions for any given Q, both P and P' are confirmed, there is no *formal* difference between the "theory" and the "assumption." However, the theory is the proposition P, which occurs in each and every hypothesis unchanged, whereas the assumption proposition is different for each H.

The actual technical procedures (i.e., the nature of such a rating operation) need not concern us in this exemplification of the formal aspects of the investigation. We would then have the following:

P_1', i.e., the demand (wish) for independence is differentially stronger in Indian college males; (P_1')

P_2', i.e., the intensity ratings for daydreams are a function of differential demand (wish) strength; (P_2')

$$\text{HYPOTHESIS: } P \cdot P_1' \cdot P_2' \supset Q_1 \qquad (H_1)$$

where Q_1 is a proposition asserting significant difference in intensity ratings; that is, if it is the case that daydreams are demand-relevant (i.e., wish-fulfilling), and in consequence of the known differential demand strength already established, and in further consequence of the functional relation between demand strength and intensity rating, it is theoretically required that the college males in the specified sample report significantly greater intensity ratings than do females.

The same formal procedures would lead to hypotheses about behavioral events which are not yet known, or at least not reported by Ruenaufer. If we consider the *frequency* with which this daydream type is reported, we should state as an hypothesis the following:

$$\text{HYPOTHESIS: } P \cdot P_1' \supset Q_1', \qquad (H_2)$$

that is, if it is the case that daydreams are demand-relevant (i.e., wish-fulfilling), and in consequence of the known differential demand strength, it is theoretically required that this daydream type be experienced with a significantly greater frequency by the male Indian college students.

While the confirmation conditions for this hypothesis are unfortunately not reported by Ruenaufer, there is some hint that the original data might well confirm the hypothesis in his statement that "even" 50 per cent of the females report it, for this certainly suggests that a good many *more* of the males did.

SUMMARY

This paper attempts an operational analysis of Freud's theory of daydreams, and it is contended that such an analysis is an indispensable prerequisite to any investigation aimed at empirical confirmation of the theory. The concept of the wish is defined in terms of an expanded version of the operation defining Tolman's concept of "demand." Formally, the procedure is hypothetical-deductive in the sense that specific hypotheses for which confirmation or unconfirmation in any investigation would be sought, are derivable from the theory, together with certain other propositions. Like all such investigations of an empirical character, however, any investigation of the theory would take the form of an inductive leap from $(P \supset Q) \cdot Q$ to (P).

The paper also discusses "levels" of theory and makes a distinction in levels between the "clinical" and "learning" language systems, with special reference to the problem of motivation.

BIBLIOGRAPHY

1. BERGMANN, G., & SPENCE, K. W. The logic of psychophysical measurement. *Psychol. Rev.*, 1944, **51**, 1–24.
2. BOSE, G. Opposite fantasies in the release of repression. *Indian J. Psychol.*, 1935, **10**, 29–41.
3. CONKLIN, E. S. Foster child fantasy. *Univ. Oregon Publ.*, 1920, **1**, No. 4.
4. CONKLIN, E. S. *Principles of adolescent psychology.* New York: Holt, 1935.
5. DEXTER, EMILY. Relation of imagination to certain other factors. *J. gen. Psychol.*, 1943, **28**, 133–141.
6. EIDELBERG, L. Contribution to the study of masturbation fantasy. *Int. J. Psycho-Anal.*, 1945, **26**, 127–137.
7. FERENCZI, S. Gulliver fantasies. *Int. J. Psycho-Anal.*, 1928, **9**, 283–300.
8. FREUD, S. *Autobiographical study.* London: Hogarth, 1936.
9. FREUD, S. Der Dichter und das Phantasieren. In *Gesammelte Werke*, London: Hogarth, 1941. Vol. VII.
10. FREUD, S. *General introduction to psychoanalysis.* Garden City: Garden City Press, 1920.
11. FREUD, S. Hysterical phantasies and bisexuality. In *Collected papers.* London: Hogarth, 1948. Vol. II.
12. FREUD, S. Hysterische Phantasien und ihre Beziehung zur Bisexualität. In *Gesammelte Werke.* London: Hogarth, 1941. Vol. VII.
13. FREUD, S. Interpretation of dreams. In *The basic writings of Sigmund Freud.* (Ed. by A. A. Brill.) New York: Modern Library, 1938.
14. FREUD, S. *New introductory lectures on psychoanalysis.* New York: Norton, 1933.
15. FREUD, S. Poet and daydreaming. In *Collected papers.* London: Hogarth, 1948. Vol. II.
16. FRIEDMANN, A. Über das Minderwertigskeitgefühl phantasievoller Kinder. *Z. pädag. Psychol.*, 1932, **33**, 273–279.
17. FÜLOP, S. The relationship between fantasy and lying in relation to the body structure of the mentally inferior. *Magyar Psychol. Szemele*, 1929, **2**, 58–74. (*Psychol. Abstr.* 1931, **5**, 1024.)
18. Green, G. H. *Psychoanalysis in the classroom.* New York: Putnam, 1922.
19. Hart, B. *Psychology of insanity.* New York: Macmillan, 1931.
20. Hesnard, A. Contribution à l'étude des phantasmes érotiques. *Rev. franc. Psychanal.*, 1930–31, **4**, 525–548.
21. HULL, C. L. Behavior postulates and corollaries. *Psychol. Rev.*, 1950, **57**, 173–180.
22. HULL, C. L. Mind, mechanism, and adaptive behavior. *Psychol. Rev.*, 1937, **44**, 1–32.
23. HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
24. HURLOCK, ELIZABETH B., & BURNSTEIN, H. Imaginary playmate: A questionnaire study. *J. genet. Psychol.*, 1932, **41**, 380–392.
25. JAEHNER, D. Die Phantasieschwester.

Z. pädag. Psychol., 1930, **31**, 371–379.

26. Kamiat, L. Cosmic phantasy. *Psychoanal. Rev.*, 1928, **15**, 210–219.
27. Koch, S. Logical character of the motivation concept. *Psychol.Rev.*, 1941, **48**, 127–154.
28. Lecky, P. *Self-consistency: A theory of personality.* New York: Island Press, 1945.
29. Lehrman, P. Significance of fantasy in neurotic behavior. *Med. J. Rec.*, 1927, **126**, 342–344.
30. Lévy-Valensi, J. Bovarysme et constitutions mentales. *J. de Psychol.*, 1930, **27**, 289–299.
31. Miller, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.
32. Mowrer, O. H. Stimulus response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, **46**, 553–564.
33. Mowrer, O. H., & Kluckhohn, C. Dynamic theory of personality. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944. Ch. 3.
34. Mowrer, O. H., & Lamoreaux, R. R. Fear as an intervening variable in avoidance conditioning. *J. comp. Psychol.*, 1946, **39**, 29–50.
35. Mowrer, O. H., & Ullman, A. D. Time as a determinant in integrative learning. *Psychol. Rev.*, 1945, **52**, 61–90.
36. Murray, H. A. *et al. Explorations in personality.* New York: Oxford Univ. Press, 1938.
37. Rogers, C. Comprehensive theory of personality. Univ. of Chicago, n.d. (Tentative draft.)
38. Rosenzweig, S. Fantasy in psychology. *J. abnorm. soc. Psychol.*, 1942, **37**, 40–51.
39. Rosenzweig, S. Systematic intent of clinical psychology. *J. abnorm. soc. Psychol.*, 1949, **44**, 3–6.
40. Ruenaufer, H. Wachträume indischer Jugendlicher. *Z. angew. Psychol.*, 1940, **59**, 319–350.
41. Shaffer, L.F. *The psychology of adjustment.* Boston: Houghton Mifflin, 1936.
42. Skinner, B. F. *Behavior of organisms.* New York: Appleton-Century, 1938.
43. Smith, Theodate. Psychology of daydreams. *Amer. J. Psychol.*, 1904, **15**, 465–488.
44. Spence, K. W. Postulates and methods of "behaviorism." *Psychol. Rev.*, 1948, **55**, 67–78.
45. Tolman, E. C. *Purposive behavior in animals and men.* New York: Century, 1932.
46. Tsai, C. Relative strength of sex and hunger motives in the albino rat. *J. comp. Psychol.*, 1925, **5**, 407–415.
47. Van Waters, M. "I would rather die than go home!" *Survey*, 1927, **57**, 565–569.
48. Varendonck, J. *Psychology of daydreams.* New York: Macmillan, 1921.
49. Vostrovsky, C. Study of imaginary companions. *Education*, 1894, **15**, 383–398.
50. Warden, C. J. *Animal motivation studies.* New York: Columbia Univ. Press, 1931.
51. White, R. W. Interpretation of imaginative productions. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944. Ch. 6.
52. Whitehead, A. N. & Russell, B. *Principia mathematica.* Cambridge: Cambridge Univ. Press, 1925.
53. Woodworth, R. S. *Psychology.* New York: Holt, 1921.

Received November 23, 1950.

THE HISTORY OF THE LEADERLESS GROUP DISCUSSION TECHNIQUE

H. L. ANSBACHER
University of Vermont

The leaderless group discussion is a technique of personality assessment in which a small group are asked to discuss a topic of common interest and are rated individually for various traits by several observers on the basis of discussion behavior. The technique is also known under the names of "group oral performance test" (**14**), "unsupervised group discussion" (**6**), and recently "group interview test" (**10**, **17**).

Since its first use in Anglo-Saxon countries during the second World War, the technique has rapidly gained in importance. Bass (**3**) gives a survey of its spread in England, Australia, and the United States. It is known to have been used for the selection of officer candidates, special military personnel, management trainees and public health officers (**3**), also of supervisors of special school classes (**10**), foremen in shipyards (**15**), and top-level civil servants (**21**). From Fields (**10**) we learn that as of 1949, at least ten public and business organizations in the United States employed the technique or experimented with it. Yet its acceptance must have reached further, since Meyer (**17**) speaks of "the untempered enthusiasm with which the group oral has been accepted in a number of places," and sees the need for warning that "there are no panaceas in the personnel field."

Validity and Reliability of the Technique[1]

The technique has gained such wide acceptance so rapidly apparently because it possesses excellent "face" validity and "logical" validity in the sense in which Cronbach (**9**, pp. 47–55) uses these terms. According to Mandell, "there is overwhelming evidence in regard to the 'face validity' of the group oral as compared with the individual oral." He quotes from three different sources. " 'The group oral examination method was highly praised . . . The consensus of the participants was that the test was fair to all, interesting.' . . . 'Feelings of confidence in the final choice have on many occasions been expressed by those who have sat on these boards. The reaction of the candidates, also, has been very encouraging.' . . . 'While this technique is undoubtedly in need of additional testing, we believe that it is superior to the individual interviews used in the past' " (**15**, p. 183).

---

[1] The writer is indebted to B. M. Bass, of Louisiana State University, for much of the information in this section.

As to empirical, correlation validity, few studies are available to date. Mandell (**15**) found a correlation of .43 between leaderless group discussion (LGD) results and a supervisory judgment test in a study of 84 foremen in two U.S. government shipyards. The correlation with ratings by colleagues and supervisors was .29, which Mandell considers a "substantial validity if one keeps in mind that the reliability of the criterion was probably not higher than .65." Bass and White (**4**) obtained correlations from .25 to .60 between LGD results and "buddy ratings" of fraternity members. This criterion had a reliability of over .90, and biserial validity correlations were used. Vernon (**21**) reports an average correlation of .36 between LGD ratings and various measures of performance of top-level civil servants in the British foreign service obtained as much as two years later. The correlations between LGD ratings and a battery of aptitude and achievement tests, heavily loaded with verbal components, averaged around .30. The LGD ratings accounted for over 50 per cent of the variance in the final disposition of the candidates, where assessment and assignment were based on written qualifying examinations, aptitude and achievement test batteries, short talks delivered by candidates, interviews, and pooled committee ratings.

Regarding reliability, only one study has been published which might be mentioned here. Bell and French (**5**) had 25 students participate in six five-man discussion groups arranged so that each individual met with completely different discussants each session. At the end of each session, the members of a group ranked their four fellow-members in order of preference for discussion leader. The average rank of the five members was then correlated with their average ranks in the other five groups in which they participated. This correlation was found to be .75 on the average.

But research concerning the LGD has only barely begun. Under B. M. Bass, of Louisiana State University, numerous studies on the subject are in progress. Work is also being done at the University of Rochester under L. F. Carter, and at Wayne University under E. T. Raney.

Origin of the Technique

The question of origin arises in connection with a new technique which has within a few years caught the imagination of many. According to accounts given in recent papers (**3**, **10**), it originated in the

British Army during World War II, as part of a much larger program of personnel assessment through group situation tests and other procedures. Regarding group situation tests in general, credit is given to the German Army (**3**), where, about 1925, such tests were actually used first. But it is not realized that the LGD, specifically, was an integral part of German military personnel selection from the beginning (**1**, p. 384). The Germans called it *Schlusskolloquium*, because the discussion took place at the end of the entire selection procedure. Sometimes they referred to it as *Rundgespräch*, which means almost literally round-table discussion.

The originator of the technique must be considered to be J. B. Rieffert, who directed German military psychology from 1920 until 1931 (**19**). Subsequently, he became an industrial psychologist, and never published anything but a few minor papers. He called the LGD simply *Kolloquium*, and sometimes conducted it over the dinner table. Rieffert states: "The round-table discussion as a selection procedure was used in Germany for the first time around 1925 . . . It was introduced by me in connection with the testing of officer candidates of the postwar [World War I] Army; I developed it together with my co-workers . . . Hans Friedländer, later *Dozent* in Berlin, now professor in England; Johannes Rudert, later professor in Leipzig; and Philipp Lersch, later professor in Munich."[2] According to Rieffert the group discussion was one of a group of four procedures used to obtain samples of problem-solving and social behavior:

1. The command series (*Befehlsreihe*), in which the candidate was asked to carry out complicated orders, shows behavior "toward specific tasks under difficult circumstances."
2. The leadership test (*Führerprobe*), in which the candidate instructed and supervised a group of men in some tasks, shows behavior "toward inferiors."
3. The group discussion (*Rundgespräch*) shows behavior "toward equal partners."
4. The interview (*Exploration*) shows behavior "toward the intellectually superior."

On these behaviors Rieffert based deductions regarding the candidate's attitudes toward work, toward his fellow-men, and toward himself.

Use of the Technique in Germany

Use in the German Armed Forces. Around 1935 German military psychology was split into Army, Navy, and Air Force psychology. The Air Force, according to Fitts (**12**, p. 155) included a group discussion in the testing of its officer candidates, but we do not know what importance was attached to this procedure. From the Army we have the following description of its use by S. Gerathewohl, a former Army psychologist, now in this country.

---

[2] Personal communication.

> The round-table discussion was already in use for the selection of officer candidates in 1934 when I became an Army psychologist . . . . After the individual interview, the entire group gathered for a final, round-table discussion (in most cases the group actually sitting around a table or in a circle). In Dresden and Breslau where I worked, the discussion was usually started by the psychologist as follows: "Now, that you are finished with all the tests, we shall talk for a while about your experiences and feelings regarding the examination. You can discuss the matter very frankly and express your opinion and criticism without being afraid that this would be to your disadvantage." Thus starting with a topic with which the candidates were very much involved and strengthening the courage of the individual by a social situation, the discussion usually became at once very lively. It was easy to shift from the initial topic of the discussion to such other topics as: officer qualifications, the importance of a strong army, etc. The directive role of the psychologist depended on the spirit of the group, the temperament displayed, and on the general level of the discussion . . . . In most cases the discussion was ended with the election of a group leader who—it was assumed—was to represent the group in an affair of honor. This democratic act was performed in the German Army as late as 1939, at least at Assessment Station IV, Dresden, under Lucke.[3]

It seems that gradually the group discussion was neglected by the Army; a search of the *Wehrpsychologische Mitteilungen*, the house organ of Army psychology, published in monthly issues of 50 to 90 pages from 1939 to 1942, did not yield any article on the technique.

In the Navy the situation was different. According to Mierke, who was in charge of German Naval psychology:

> The group discussion gained continuously in importance with us and came to be one of the weightiest procedures . . . . It remained a part of the abbreviated officer selection program of the late years of the war and was used occasionally even in the selection of specialists.[4]

Around 1938 Mierke had modified the discussion into an openly controversial one because he had found this more revealing.

> The form of defense and attack; the striving for objectivity, for conciliation or for compromise solutions; . . . sense of humor and the lack of it; slyness and rudeness—in short almost all forms of behavior in social relations can become manifest in this procedure. As a rule it will even be necessary to calm heated tempers at the conclusion by asking that the entire affair be regarded only as an amusing intellectual game and that the clash of opinions be not continued elsewhere (**18**, p. 62).

Controversial topics generally suitable for 18-year-old youths were found to be "For or Against Dancing Lessons" or "Moderate Smoking versus No Smoking." Topics of a military or political nature were avoided on purpose to obtain spontaneous rather than drilled or dishonest behavior.

---

[3] Personal communication.

[4] Personal communication.

Civilian use in wartime Germany. With the growing manpower shortage in wartime Germany, psychology experienced a boom in that it was increasingly applied in the nationwide talent searches which were conducted in many fields. The actual selection of talent frequently took the form of selection camps, modeled after the tried military method (**2**, pp. 610–611). Among the procedures used was the group discussion.

A detailed recommendation of procedure for use in metal-industry camps for the selection of able young workers is to be found in an article by Fischer and Lottmann (**11**). The camp session, which is to last about ten days, would include three group discussions of about one hour each. The first discussion would refer to occupational goals, the second to some technical problem, the third to impressions gained from a preceding visit to a nearby plant.

Dr. Christel Drey-Fuchs, who worked as a psychologist in trade, commercial, educational, and artistic selection camps, states that one or more group discussions were generally carried out.[5] There is also a corroborative account by Dr. Helga Schmidt-Oesfeld, a psychologist who had been examined in such a camp in connection with her application for a study grant.[6]

Use of the technique in present-day Germany. Today the group discussion is used in Germany as part of the entrance examination to some teachers colleges. Stückrath (**20**) conducted an informal survey among teacher-training institutions and found that the group discussion is favored "where for lack of time one is compelled to adhere to the more conventional methods of selection and yet does not want to do without a systematic observation of behavior." Among the topics are: "Are Wars Avoidable?" "Modern Dance," "Alcohol and Nicotine," "Coeducation," and "Corporal Punishment." Factors to be considered in the evaluation are: Is the examinee in rapport with the group? Does he respect the opinion of others? Does he give the other fellow a chance to talk? In all, the group discussion is seen as a test of social attitudes.

Specifically, the group discussion is used at the Teachers College (*Pädagogische Hochschule*) in Kiel under Mierke,[7] the former director

---

[5] Personal communication.

[6] Personal communication.

[7] Personal communication.

Bureau Ednl. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated

of Naval psychology, who considers one of the merits of the procedure to be that nonpsychologists can become reasonably well trained in its application. Another teachers college where the group discussion is used in the selection of students is the Pedagogical Institute of Darmstadt at Jugenheim. An American exchange professor, who had a chance to observe these discussions, stated that he was greatly impressed by them and their possibilities and that they were superbly handled by Professor Ruppert. In an interview with the present writer Ruppert explained that the group discussion is used for the appraisal of personality and social attitudes. His preferred topics are "The Ideal Teacher" and "What Difficulties Do I Expect to Encounter in Teaching?" He found it advantageous to seat the candidates in an open circle—rather than around a table—because when deprived of the prop of a table, they are literally and figuratively speaking more open to inspection. In addition to about eight candidates, two to three judges are included in the circle, and these have a table in front of them for note-taking. A judge would occasionally interfere to help a particularly shy individual enter the discussion. Ruppert had been an Army psychologist at one time during the war.[8]

Social-Psychological Implications

We have shown that the group discussion test was introduced in Germany twenty-five years ago and has been holding its place there ever since. While the details of the picture have not been presented heretofore, the basic fact had been mentioned in the American psychological literature at least four times—in 1941 (**1**, p. 384), in 1942 (**16**, p. 176), in 1945 (**22**), and in 1946 (**12**, p. 155)—in media of widest circulation and often-quoted references, with exception of the 1945 reference, which is only an abstract. The paper by Martin (**16**, p. 176) devoted half a page to a description of the technique.

How is it to be explained that American psychologists interested in the group discussion have nevertheless remained oblivious to the fact that it was originally used in Germany? One explanation, of course, might be that it was simply overlooked. But might it not also be that the group discussion method, with its aspects of spontaneity, fluidity, and democracy, did least fit into the prevailing simplified conception of the German scene, particularly as it appeared under the Hitler dictatorship? In that case we would have an example in the field of social psychology of the frequently noted mechanism of maintaining an original conception by overlooking elements which would disturb it.

---

[8] Since this paper was completed, it has become known that the group discussion is being used for the assessment of officer candidates in the newly created West-German emergency police. (Scharmann, T. Bericht über psychologische Untersuchungen bei den Auswahllehrgängen der Bereitschaftspolizei in Traunstein. *Psychol. Rundschau*, 1951, **2**, 115.

A paper by Jennings (13) gives an eloquent illustration of the mechanism by which stereotypes are protected. She states:

> As we study the situation tests used by the military psychologists in Germany under the Nazi regime . . . we note that not one of them allows the individual scope and variety in solutions, nor gives him a chance for personality expression *per se* . . . . He is never placed in a setting where he has opportunity *to develop a relationship between himself and specific other persons* . . . nor in any situations specifically constructed *to be meaningful to him as a particular individual* . . . . Under the Nazi regime, it may very well be that the psychological climate could ill afford to encourage spontaneous expression either in its experimental program or its regime as a whole. (13, p. 191).

Jennings sees complete consistency between psychological methods and political climate.

Actually, however, any social phenomenon is complex and very likely to include inconsistencies. Recently, an entire symposium was devoted to the problem of inconsistency in the social attitudes and the behavior of one and the same individual (8). One of the conclusions reached by Chein was that "The first step toward improving the quality of research in intergroup relations is the awareness of consistency-inconsistency as a pertinent dimension" (7, p. 52). If inconsistency regarding social behaviors is such an important factor on the individual-psychological level, how much more must we be prepared for inconsistencies on the institutional-sociological level of an entire nation. The discovery of an existing inconsistency, while disturbing to stereotypes, is highly welcome to those interested in changing a given attitude or condition, because the inconsistency affords a basis for any attempt to bring about change (7, p. 59).

The discovery of an inconsistency also presents a challenge for closer examination. With regard to the present situation we find that the Nazi regime was not as completely totalitarian as it appeared. Military psychology was not particularly Nazi, and relatively few psychologists were members of Nazi organizations. That is not to say that those in charge were not solid militarists; but the minor psychologists were not necessarily even that. Actually, the type of people military psychology attracted earned it the nickname of "internal emigration." Its partial dissolution during the war may well have been caused by the smoldering conflict with the Nazi party and principles, as many of these psychologists have claimed. In fact, the Air Force, which stood most strongly under Nazi influence, dismissed its psychologists first; next came the Army. The Navy, which remained most autonomous, retained its psychologists to the end of the war. The Navy was also the most faithful user of the LGD.

Summary

The leaderless group discussion test has received increasing attention in Anglo-Saxon countries since World War II through its very satisfactory face and logical validity. Available correlational research regarding the technique is still scanty, but a considerable amount of such work is in progress. The importance which the technique has achieved has led the author to an outline of its history in Germany from its beginnings in 1925 to the present day. It was used in the German Army and Air Force, but particularly in the German Navy; it was used in Nazi-sponsored selection camps devoted to a widespread talent search; it is used in postwar Germany as part of the entrance examination to certain teachers colleges. The use of such an unstructured, democratic technique does not fit the simplified conception of Germany which still prevails, and this may be one reason why the German priority in the group discussion has heretofore been overlooked.

BIBLIOGRAPHY

1. ANSBACHER, H. L. German military psychology. *Psychol. Bull.*, 1941, **38**, 370–392.
2. ANSBACHER, H. L. German industrial psychology in the fifth year of war. *Psychol. Bull.*, 1944, **41**, 605–614.
3. BASS, B. M. The leaderless group discussion technique. *Personnel Psychol.*, 1950, **3**, 17–32.
4. BASS, B. M., & WHITE, O. L. Validity of leaderless group discussion observers' descriptive and evaluative ratings for the assessment of personality and leadership status. *Amer. Psychologist*, 1950, **5**, 311–312. (Abstract.)
5. BELL, B., & FRENCH, R. L. Consistency of individual leadership position in small groups of varying membership. *J. abnorm. soc. Psychol.*, 1950, **45**, 764–767.
6. BRODY, W. Judging candidates by observing them in unsupervised group discussion. *Personnel J.*, 1947, **26**, 170–173.
7. CHEIN. I. The problem of inconsistency: A restatement. *J. soc. Issues*, 1949, **5**, No. 3, 52–61.
8. CHEIN, I., *et al.* (Eds.) Consistency and inconsistency in intergroup relations. *J. soc. Issues*, 1949, **5**, No. 3.
9. CRONBACH, L. J. *Essentials of psychological testing.* New York: Harper, 1949.
10. FIELDS, H. The group interview test: Its strength. *Publ. Personnel Rev.*, 1950, **11**, 139–146.
11. FISCHER, G. H., & LOTTMANN, W. Menschenbild, Auslese und Erziehung. Beitrag VIII: Zur Methodik im beruflichen Ausleselager. *Z. pädag. Psychol.*, 1943, **44**, No. 5, 85–93.

12. Fitts, P. M. German applied psychology during World War II. *Amer. Psychologist*, 1946, **1**, 151–161.
13. Jennings, Helen H. Military use of sociometric and situation tests in Great Britain, France, Germany and the United States. *Sociometry*, 1949, **12**, 191–201.
14. Mandell, M. M. The group oral performance test. *Publ. Personnel Rev.*, 1946, **7**, 209–212.
15. Mandell, M. Validation of group oral performance test. *Personnel Psychol.*, 1950, **3**, 179–185.
16. Martin, A. H. Examination of applicants for commissioned rank. In *L. Farago (Ed.), German psychological warfare* (1st definitive Ed.) New York: Putnam, 1942. Pp. 171–178.
17. Meyer, C. A. The group interview test: Its weakness. *Publ. Personnel Rev.*, 1950, **11**, 147–154.
18. Mierke, K. Psychologische Diagnostik. In N. K. Ach (Ed.), *Lehrbuch der Psychologie: Praktische Psychologie.* Bamberg: C. C. Buchner, 1944. Vol. 3, pp. 1–79.
19. Simoneit, M. Vom Werden der deutschen Wehrmachtpsychologie; ein geschichtlicher Rückblick. *Wehrpsychol. Mitt.*, 1940, **2**, Suppl. No. 2.
20. Stückrath, F. Der gegenwärtige Stand der Ausleseverfahren für den Lehrerberuf. *Bildung und Erziehung*, 1949, **2**, 501–516.
21. Vernon, P. E. The validation of Civil Service Selection Board procedures. *Occup. Psychol., Lond.* 1950, **24**, 75–95.
22. Winzer, H. A. Abstract of Fischer, G. H., & Lottmann, W. Menschenbild, Auslese und Erziehung. Beitrag VIII. Zur Methodik im beruflichen Ausleselager. *Z. pädag. Psychol.*, 1943, **44**, No. 5, 85–93. (*Psychol. Abstr.*, 1945, **19**, No. 3465.)

Received January 20, 1951.

COMPUTATION OF THE LEVEL OF SIGNIFICANCE IN THE *F*-TEST

C. J. BURKE
Indiana University

Frequently, the psychological experimenter expresses dissatisfaction with the ordinary table for interpreting *F*-ratios. This table gives the values of the *F*-ratio which are required for significance at the .01 and .05 levels, but, unlike the t and χ^2 tables, interpolations for values below the .05 level are not possible. Faced with a value of F which is not significant at the .05 level, the experimenter often wishes to estimate its actual significance. If it is significant at the 8 per cent level, for example, further experimentation may be indicated, but if it is barely significant at the 30 per cent level, the experimenter might discontinue working with the particular variables involved.

One way of resolving this state of affairs would be the provision of an adequate *F*-table. The *F*-table is essentially three-dimensional and cannot be reduced to two dimensions—its adequate representation would require a book, rather than a page. In spite of the extensive labors that would be required, it might be of use to tabulate the entire *F*-distribution for such numbers of the degrees of freedom as occur rather frequently, were it not for the fact that an extensive table of an intimately related distribution already exists. Pearson (**2**) has tabulated the beta-distribution, and the book in which this distribution is represented is rather widely available in university libraries. A simple computational procedure enables one to assign a level of significance to any value of F by means of the tabulated beta-distribution.

The intimate relation between these two distributions has long been known to professional statisticians, and the underlying theory relating them is presented in rather widely used treatises by Cramèr (**1**), Wilks (**3**), and many others. In the present paper, the results of this theory will be summarized in a form which is mathematically less formidable than the usual presentations. Following the presentation of the theory, the detailed steps in the calculations are illustrated by means of several examples. The psychological statistician who is faced with the problem of calculating the level of significance of a value of F and who is not interested in the underlying theory can use the examples as a model for his calculations without reading the theoretical section.

THEORY

With the quantity which we compute as F is associated a probability distribution from which the probability that a value of F selected at random lies between any two given values can be calculated. We shall symbolize this distribution in its cumulative form by $G_{r_1, r_2}(F)$, where r_1 and r_2 are the number of degrees of freedom respectively associated with the variances in the numerator and denominator of the F-ratio. (this means that r_1 will actually occur in the denominator and r_2 in the numerator of the F-ratio, since that ratio is obtained by dividing variance estimates). If we are given a specific value of F, denoted by F_0, with specified values of r_1 and r_2, the probability that any value of F selected at random will be smaller than F_0 can be calculated from:

$$P(F \leqq F_0) = G_{r_1,r_2}(F_0). \qquad [1]$$

The ordinary F-table gives the values of F_0 for each combination of r_1 and r_2 values which makes the probability equal to .99 or .95. If we wish to work at levels of confidence other than .01 or .05, the table is of little use.

Because the basic definition of F involves the ratio between two values of chi-square, the reciprocal of F also has the F distribution, but with the order of r_1 and r_2 reversed. Hence,

$$P\left(\frac{1}{F} \leqq \frac{1}{F_0}\right) = G_{r_2,r_1}\left(\frac{1}{F_0}\right). \qquad [2]$$

But, from the nature of reciprocals,

$$P\left(\frac{1}{F} \leqq \frac{1}{F_0}\right) = P(F \geqq F_0) = G_{r_2,r_1}\left(\frac{1}{F_0}\right). \qquad [3]$$

Also, since the distribution is continuous,

$$P(F \geqq F_0) = P(F > F_0) = 1 - P(F \leqq F_0). \qquad [4]$$

From [1] and [4],

$$P(F \geqq F_0) = 1 - G_{r_1,r_2}(F_0). \qquad [5]$$

The combination of [3] and [5] yields a relation between the two cumulative distribution functions:

$$G_{r_2,r_1}\left(\frac{1}{F_0}\right) = 1 - G_{r_1,r_2}(F_0). \qquad [6]$$

Subsequent use will be made of equation [6]. It is well known (3, pp. 114–115) that there exists an intimate connection between the F-distribution and the incomplete beta-function tabulated by Pearson. If we sample at random from an F-distribution and convert each value of F obtained to a value of x according to the formula

$$x = \frac{\dfrac{r_1}{r_2} F}{1 + \dfrac{r_1}{r_2} F}, \qquad [7]$$

the values of x obtained will range between 0 and 1 and will have the distribution tabulated by Pearson. When a given value F_0 has been selected, a value of x, denoted by x_0, can be computed from

$$x_0 = \frac{\dfrac{r_1}{r_2} F_0}{1 + \dfrac{r_1}{r_2} F_0}. \qquad [7']$$

Then

$$P(x \leqq x_0) = P(F \leqq F_0). \qquad [8]$$

The probability on the left of [8] can be obtained from Pearson's tables. The cumulative distribution function of x is written symbolically as $I_{p,q}(x)$[1] so that

$$P(x \leqq x_0) = I_{p,q}(x_0) \qquad [9]$$

where

$$p = \frac{r_1}{2} \quad \text{and} \quad q = \frac{r_2}{2}. \qquad [10]$$

From [1], [8], and [9],

$$G_{r_1,r_2}(F_0) = I_{p,q}(x_0). \qquad [11]$$

Pearson's tables give values of $I_{p,q}(x)$ for various values of p, q, and x, such that $q \leqq p$. This restriction on values of p and q means that we must consider two cases.

Case 1. $r_1 \geqq r_2$, $(p \geqq q)$. We are given a value F_0 and seek to compute its level of significance $[1-G_{r_1,r_2}(F_0)]$. We calculate p, q, and x_0 from [10] and [7′]. Then we look up the value of $I_{p,q}(x_0)$ in Pearson's tables. This value is $G_{r_1,r_2}(F_0)$ according to [11] and hence we obtain the desired level of confidence by subtracting this value from unity.

---

[1] The notation here employed differs slightly from that used by Pearson, but the correspondence between the two notational systems is obvious and should cause no confusion.

Case 2. $r_1 < r_2$, $(p < q)$. Again, we are given a value of F_0 and desire the value of $1-G_{r_1,r_2}(F_0)$. If we follow the method outlined in case 1, we are unable to find $I_{p,q}(x_0)$ in Pearson's table since values for $p<q$ are not tabulated. To get around this difficulty, we make use of the fact that the reciprocal of F also has the F distribution. We calculate x_0 according to:

$$x_0 = \frac{\dfrac{r_2}{r_1}\dfrac{1}{F_0}}{1+\dfrac{r_2}{r_1}\dfrac{1}{F_0}}. \qquad [12]$$

For the reader who is familiar with the beta-function, the significance of the change from formula [7′] to formula [12] will be clarified by noting that, if x is calculated from formula [7′], the quantity on the right of [12] gives $1-x$. Thus, changing r_1 and r_2 in the F-distribution changes p and q in the beta-distribution.
Then

$$G_{r_2,r_1}\left(\frac{1}{F_0}\right) = I_{p,q}(x_0) \qquad [13]$$

where

$$p = \frac{r_2}{2} \quad \text{and} \quad q = \frac{r_1}{2}\cdot \qquad [14]$$

With equation [14], $p \geqq q$, and the value of $I_{p,q}(x)$ can be found in Pearson's table. From [13] and [6], we have

$$1 - G_{r_1,r_2}(F_0) = I_{p,q}(x_0) \qquad [15]$$

so that the value of $I_{p,q}(x_0)$ is taken directly as the level of significance.

COMPUTATIONAL EXAMPLES

Case 1. $r_1 > r_2$. *Example 1.* We are given a value of $F_0 = 1.30$ with degrees of freedom $r_1 = 30$ and $r_2 = 19$ associated with the numerator and denominator, respectively. According to [7′], we calculate x_0 as

$$x_0 = \frac{\dfrac{30}{19}(1.30)}{1+\dfrac{30}{19}(1.30)} = .673,$$

and $p=15$, $q=9.5$ from [10]. From Pearson (2, p. 240), we obtain

$$I_{15,9.5}(.670) = .713$$
$$I_{15,9.5}(.680) = .748.$$

By interpolation

$$I_{15,9.5}(.673) = .724.$$

Hence the level of significance is $1-.724=.276$.

Example 2. Given $F_0=1.50$, $r_1=27$, $r_2=25$, find the level of significance.

$$x_0 = \frac{\frac{27}{25}(1.50)}{1+\frac{27}{25}(1.50)} = .618$$

from [7′], and $p=13.5$, $q=12.5$ from [10]. From Pearson, we obtain on page 278 ($q=12$)

$$I_{13,12}(.610) = .815 \qquad I_{14,12}(.610) = .765$$
$$I_{13,12}(.620) = .842 \qquad I_{14,12}(.620) = .796$$

and on page 288 ($q=13$)

$$I_{13,13}(.610) = .870 \qquad I_{14,13}(.610) = .829$$
$$I_{13,13}(.620) = .891 \qquad I_{14,13}(.620) = .855.$$

At this juncture, interpolation with respect to three variables is necessary. To obtain really accurate results, higher-order interpolation formulas should be used, but for most purposes, linear interpolation will yield sufficient accuracy. Interpolating first with respect to q, we obtain

$$I_{13,12.5}(.610) = .842 \qquad I_{14,12.5}(.610) = .797$$
$$I_{13,12.5}(.620) = .866 \qquad I_{14,12.5}(.620) = .826.$$

Interpolation with respect to p yields

$$I_{13.5,12.5}(.610) = .820$$
$$I_{13.5,12.5}(.620) = .846.$$

It should be noted that the interpolations with respect to p and q can be done in one step:

$$I_{13.5,12.5}(.610) = \frac{.815+.870+.765+.829}{4} = \frac{3.279}{4} = .820$$
$$I_{13.5,12.5}(.620) = \frac{.842+.796+.891+.855}{4} = \frac{3.384}{4} = .846.$$

Finally, we interpolate with respect to x_0 to obtain

$$I_{13.5,12.5}(.618) = .841.$$

Hence, the level of significance is $1-.841=.159$.

Case 2. $r_1<r_2$. *Example 3.* We are given a value of $F_0=1.50$ with $r_1=25$ and $r_2=75$. Using equation [12], we have

$$x_0 = \frac{\frac{75}{25}\left(\frac{1}{1.50}\right)}{1+\frac{75}{25}\left(\frac{1}{1.50}\right)} = .667$$

and from [14], $p=37.5$, $q=12.5$. From Pearson's tables on page 285 ($q=12$), we obtain

$$I_{37,12}(.660) = .0678 \qquad I_{38,12}(.660) = .0564$$
$$I_{37,12}(.670) = .0886 \qquad I_{38,12}(.670) = .0748$$

and on page 294 ($q=13$), we obtain

$$I_{37,13}(.660) = .1027 \qquad I_{38,13}(.660) = .0870$$
$$I_{37,13}(.670) = .1313 \qquad I_{38,13}(.670) = .1126.$$

Interpolation with respect to p and q yields:

$$I_{37.5,12.5}(.660) = \frac{.0678 + .0564 + .1027 + .0870}{4} = .0785$$

$$I_{37.5,12.5}(.670) = \frac{.0886 + .0748 + .1313 + .1126}{4} = .1018.$$

Finally, interpolation with respect to x_0 gives

$$I_{37.5,12.5}(.667) = .0948.$$

In case 2, no subtraction is necessary as this figure gives the level of significance directly.

SUMMARY

In this paper, the well-known theory by means of which the significance of any obtained F-ratio can be obtained from tables of the incomplete beta-function is summarized. Computational examples are presented which can serve as models for investigators interested in determining the exact significance of values which do not occur in the ordinary table for F.

BIBLIOGRAPHY

1. Cramèr, H. *Mathematical methods of statistics*. Princeton: Princeton Univ. Press, 1946.
2. Pearson, K. *Tables of the incomplete beta-function*. London: *Biometrika* office, 1934.
3. Wilks, S. S. *Mathematical statistics*. Princeton: Princeton Univ. Press, 1944.

Received November 22, 1950.

ON THE USE OF LATIN SQUARES IN PSYCHOLOGY

QUINN McNEMAR
Stanford University

The purpose of this note is to attempt a clarification of two aspects of the latin square design which seem inadequately treated in the expositions written primarily (**2, 4, 5, 11**) or partly (**6, 10**) for psychologists. First we shall consider briefly the types of situations where the latin square design might be useful, and second we shall point out the statistical assumptions which must be met. Then we shall face the implications of the latter for the former.

To facilitate the discussion, suppose the agricultural situation represented by the accompanying square, in which the letters stand for

		Rows			
		1	2	3	4
Columns	I	*A*	*D*	*B*	*C*
	II	*B*	*A*	*C*	*D*
	III	*C*	*B*	*D*	*A*
	IV	*D*	*C*	*A*	*B*

four different treatments so arranged in a field plot that each treatment occurs once in each row and once in each column. The object is to average out possible fertility differentials from row to row and from column to column. With soil heterogeneity thus balanced, the experiment is obviously under better control, hence the results should have greater precision; this greater precision is reflected in an error term which is the residual *after* variations due to row differences and to column differences (also treatment effects) have been deducted.

Since the design need not be earth-bound, it is only natural that psychologists should adopt it for those situations where they wish to balance out the effects of sources of variation not experimentally controllable. The letters are assigned to "treatments" (dosages, methods, conditions, etc.) while the rows and columns typically stand for subjects and order of testing. Variations due to subjects and due to order (practice or fatigue) are consequently balanced as regards the treatments, and the statistical analysis capitalizes on this fact in that the error term used for the F ratio does not include variation associated with subjects and with order.

The latin square was not invented for the purpose of providing a means for taking care of variation due to easily controlled variables, e.g., illumination. However, the design can be used for situations in which the columns and rows need not represent uncontrollable sources of variation. Either the rows or the columns or both can stand for factors which are deliberately varied for the sake of testing hypotheses regarding their effects. That the latin square design may be used instead of the so-called factorial design seems to have been overlooked in the expositions for psychologists. As early as 1937, Fisher, in his *The Design of Experiments* (**3**, section 35.1), mentioned the possibility. More recently the idea has appeared in books by Kendall (**7**), Cochran and Cox (**1**), and Mood (**9**). Stated very briefly, a latin square design with 16 properly arranged observations may be used in place of a complete three-way analysis of variance (factorial) design requiring 64 observations; 25 observations may be used instead of the 125 needed in a three-way analysis with five levels for each classification (or factor); and so on.

Obviously, each factor must involve the same number of levels to permit the substitution of a latin square design for a complete factorial design, and equally obvious is the fact that letters in such squares stand for levels on one of the factors. This use of a latin square allows the testing of hypotheses regarding all three factors. As usual, the error term is free of variation due to the three factors under study. Unlike the complete three-way analysis of variance, the interactions cannot be tested. The chief advantage of the use of a latin square in place of a factorial design is that fewer observations are required—an important consideration when the securing of an observation is costly.

It has been implied above that the latin square is permissible for a mixed situation—either the rows (or columns) standing for a factor to be investigated while the columns (or rows) stand for an uncontrollable variable the effect of which needs to be balanced. Accordingly, the latin square would seem to be a flexible design, which it would be if it were not for the matter of assumptions.

Aside from the usual assumption that the observations are from normally distributed populations with equal variances, *it is also assumed that all interactions are zero.* For some reason or other the advocates of the use of the latin square in psychological research have passed over this second and fundamental assumption. Though Grant (**5**) comes nearest to pointing out this assumption when he says that the interactions are confounded, cannot be assessed, and "may influence the size of the error term," he fails to make it explicit. One searches in vain for mention of this assumption in the 30-page chapter on latin squares in Edwards' recent book (**2**). The assumption is hinted at in Fisher's

Design (**3**, section 35.1), while it is directly implied in the mathematical formulations of Cochran and Cox (**1**, pp. 41–42), of Mann (**8**, p. 77), of Mood (**9**, p. 340), and of Wilks (**12**, p. 190). Mood also puts it in near Basic English: "all interactions are assumed to be zero."

What is the possible consequence of failure to meet this assumption? From the viewpoint of statistical theory it simply means that F's from latin squares do not follow the F distribution—too many "significant" F's will be obtained when the assumption is not met. This will happen because the residual term involves an admixture of ordinary error and any two-way interaction that is present, which interaction will of course be larger than the ordinary error component, but the combination of the two sources will tend to yield a residual which is smaller than the interaction that properly should be used as the denominator for F.

What of the likelihood that the assumption will not hold for psychological variables? First, consider the possible use of a latin square in lieu of a complete factorial design. One need only glance through the *Journal of Experimental Psychology* to learn that in about half the studies involving the testing of interactions, significant interactions between the factors emerge. Second, consider the commonly used situation in which either the rows or the columns stand for individuals. Since there is nearly always an interaction between individuals and factors, it follows that the assumption will nearly always be violated in such situations.

Faced with these facts, we are forced to the inescapable conclusion that the latin square design is seldom appropriate in psychological research. *It is defensible only in those rare instances when one has sound a priori reasons for believing that the interactions are zero.*

BIBLIOGRAPHY

1. Cochran, W. G., & Cox, G. M. *Experimental designs*. New York: Wiley, 1950.
2. Edwards, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
3. Fisher, R. A. *The design of experiments*. (2nd Ed.) Edinburgh: Oliver and Boyd, 1937.
4. Garrett, H. E. & Zubin, J. The analysis of variance in psychological research. *Psychol. Bull.*, 1943, **40**, 233–267.
5. Grant, D. A. The latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1948, **45**, 427–442.
6. Johnson, P. O. *Statistical methods in research*. New York: Prentice-Hall, 1949.
7. Kendall, M. G. *The advanced theory of statistics*. London: Griffin, 1948. Vol. II.
8. Mann, H. B. *Analysis and design of experiments*. New York: Dover, 1949.

9. Mood, A. M. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.
10. Peters, C. C., & Van Voorhis, W. R. *Statistical procedures and their mathematical bases.* New York: McGraw-Hill, 1940.
11. Thomson, G. H. The use of the latin square in designing educational experiments. *Brit. J. educ. Psychol.*, 1941, 11, 135–137.
12. Wilks, S. S. *Mathematical statistics.* Princeton: Princeton Univ. Press, 1943.

Received November 29, 1950.

THE GENETICS OF SCHIZOPHRENIA: REPLY TO PASTORE[1]

LEWIS A. HURST

Weskoppies Hospital, Pretoria, South Africa

The two main technical sources of Pastore's errors in his criticisms of Kallmann's *The Genetics of Schizophrenia* (1) are:

1. Ignorance of the more recent experimental developments in genetics and the conceptual framework of physiological genetics erected thereon.

2. Ignorance of certain special statistical techniques relevant to genetic and population studies evolved chiefly in Germany.

The acclaim accorded to Kallmann's work (**1**, **5**) by Hogben and Haldane at the International Congress of Genetics at Edinburgh in 1939, in the face of their formerly contrary preconceptions, should give one pause before accepting Pastore's picture of Kallmann as a simple-minded medico barging clumsily and unwarily into the domains of science and statistics. Hogben and Haldane are people who *do* know about modern physiological genetics and the statistical methods evolved in connection with genetic and population problems.

Pastore's attitude gains a spurious impetus from the tendency current in "advanced" circles to identify gratuitously a belief in human heredity with a reactionary attitude and that in the all-powerful influence of environment with a progressive attitude. Elsewhere (**2**, **3**), I have sought to indicate the scientifically inacceptable factors of a personal and professional nature entering into this type of environmentalism, as well as the illegitimate extension, by analogy, of Freudian concepts from hysteria to schizophrenia, and the guilelessness of the implied claim of Watson, advanced without any experimental verification, as to the possibility of making a mental defective or genius of an identical individual by implanting different sets of conditioned reflexes.

Before proceeding to a detailed reply to Pastore's criticisms of Kallmann, I feel that there is an onus on me to elaborate and clarify my allegations against him on the score of ignorance of genetics and of certain special statistical techniques evolved in relation to genetics and population studies.

The history of genetics may be divided into three stages:

1. The first is that associated with the name and work of Mendel, in which

---

[1] Pastore, N. The genetics of schizophrenia: A special review. *Psychol. Bull.*, 1949, **45**, 285–302.

the genes or hereditary units were merely hypothetical entities (whose anatomical basis was unknown) postulated to explain the ratios obtained in crossbreeding experiments.

2. The second stage, associated with the name and work of T. H. Morgan, has as its salient feature the location of Mendel's hypothetical units on a physical substrate, the chromosomes, and in certain species the construction of maps in which the genes responsible for specified traits were assigned positions relative to one another on the individual chromosome pairs.

3. The third stage, that of modern physiological genetics, is associated with the name and work of Goldschmidt. Goldschmidt's experimental developmental studies drew attention to the following fact: the biochemical processes originating from a particular gene pair or set occur in the environment of the rest of the organism, which is the result of the biochemical processes originating from all the other genes of that organism. It is clear that as a result of chemical interaction the processes initiated by the gene pair under consideration may be furthered, inhibited, or modified by the processes deriving from the other genes. Hence the recognition that the genotype, i.e., the sum total of all the component genes of the fertilized ovum, is seldom fully manifested in the phenotype, defined by Kallmann as "manifest features of an organism, representing the end product of the development and appearance of all inherited characters in the individual." Hence, also, the concept of modifying genes or modifiers; and the concept, so very important in modern genetics, of *penetrancy*, which denotes the percentage of cases in which the trait associated with a particular pair or set of genes becomes manifest in the developed organism. The obverse of this concept of penetrancy is the modern tendency in genetics to speak of genetic predisposition, which carries with it a recognition that, few traits being fully or 100 per cent penetrant, the possession by an organism of the requisite complement of genes for a given trait does not imply the invariable appearance of the trait in the fully developed organism, but appearance only in a certain percentage of cases. In the sphere of the clinical entity with which we are concerned, viz., schizophrenia, the concept of constitution (and constitutional resistance) has, as its genetic correlate, modifying genes, which reduce the penetrancy of the single recessive gene pair responsible for the hereditary predisposition to schizophrenia from 100 per cent to about 70 per cent, as Kallmann has shown.

It is clear from Pastore's article that he is living mentally in stage 1 of the history of genetics, not in stage 3, modern physiological genetics, and that is why he fails to understand Kallmann's work. This appears, *inter alia*, from his naive reference, of which he is oblivious of any need for qualification, to "Kallmann's Mendelian outlook."

As regards the statistical methods evolved in relation to human genetics, a full exposition and application to schizophrenic material is to be found in Kallmann's *The Genetics of Schizophrenia*. They are concerned with expectancy figures as opposed to net figures. Here we can only name the methods and indicate their general rationale. All the methods take cognizance of the fact of whether the relative in question has not reached, has reached, or has passed the danger period for the manifestation of the trait under consideration. In the case of schizophrenia this is taken as from the fifteenth to the forty-fourth year. In exposition of Weinberg's abridged method Kallmann says: "Accordingly, we counted only half of persons who were between the ages of fifteen and forty-four at the end of our statistical control. All persons who were forty-five or older were counted in full, while those who dropped out of our statistics before the age of fifteen were omitted entirely." The Ilse method and morbidity statistics claim greater accuracy than Weinberg's abridged method in that they do not postulate a uniform age distribution within the danger period of manifestation, but there are serious disadvantages in Ilse's method which lead Kallmann after careful analysis to rank it as inferior to Weinberg's abridged method. The detailed mathematical exegesis given by Kallmann in connection with these methods (especially p. 139[2]), and also Weinberg's double proband method, and Schulz's double case method based on Bernstein's a priori method (especially p. 147), accords ill with the simple-minded medico theory. In Weinberg's proband method the principle is adopted that "a proband can be included in the estimate only if he also appears in the survey as the sibling of a proband, while, conversely, the sisters and brothers of a proband should not be selected as probands merely because they are carriers of the schizophrenic trait." Weinberg's double proband method "employs only series of siblings with at least two probands, counting them by means of the usual proband method, as a double proband entity and reckoning their other siblings singly."

Similarly, the double case method is based on Bernstein's *a priori* method, which is used for the separate estimate of the frequency of the given trait both in the series of siblings with at least two trait carriers, and in the total sibling material. The principle of this method rests on the assumption that the hereditary quality of a supposedly transmitted trait, in general occurring rarely, may be viewed as demonstrated if it is found in several individuals of a series of siblings; while there is a stronger possibility of exogenous origin if it is manifested in only one child of a series. Accordingly, if the percentages for the frequency of an hereditary trait in the total survey agree with the figures which are obtained by the same statistical method for the series of siblings with two or more trait carriers, it is highly probable that the proband material under investigation is biologically homogeneous and the given trait is conditioned by heredity.

---

[2] The page references cited in this article are to Kallmann's, *The Genetics of Schizophrenia* (4).

Kallmann outlines the mathematical techniques evolved by Weinberg and by Schulz to deal with the factors complicating the a priori method in the case of schizophrenia; these factors are that "we are dealing with a late developing and not completely penetrant trait, and a material which comprises a collection of diseased individuals and is not selected directly from the general population." In the section "The Genetic Relation between Schizophrenia and Tuberculosis" (chap. VII), the *G* and *T* methods are differentiated. In the former the tuberculosis mortality rates are related to the totals of observed individuals, while the *T* method deals only with the sum of all deceased persons as a corrected rate of reference.

This brief survey will give some notion of the diversity and complexity of the statistical methods employed by Kallmann, of which his critic shows little if any grasp.

It it now time to answer the detailed criticisms preferred against Kallmann's work.

Establishing the Diagnosis

1. The necessary limitations of a study in human heredity designed to include adequate numbers of ancestors and descendants is appreciated as clearly by Kallmann as by his critic. Kallmann, moreover, emphasizes the need for further studies in many directions.

2. The clinical records of a leading German metropolitan mental hospital at both extremes of the quarter-century referred to were of a standard to permit satisfactory objective assessment, by a modern investigator, of the diagnoses.

3. The clinical content of schizophrenia was identical at the beginning and end of the quarter-century referred to; the only change was that of name—from dementia praecox to schizophrenia.

4. Institutionalization is by no means a necessary feature in the diagnosis of schizophrenia.

5. The exact clinical status and nature of the "doubtfuls" is clearly portrayed for those who take the trouble to read Kallmann's book (especially p. 15). Kallmann's scientific honesty is so strict that his "doubtfuls" probably contain a high proportion of schizophrenics.

6. No statistical error is introduced by Kallmann's division into "definites" and "doubtfuls"; the distinction is carefully maintained both in the tables and in the text, for the first-hand inspection of the readers, and the "doubtfuls" were not included in the calculation of final figures.

7. Whether the "doubtfuls" will become "definites" is an idle question, indicating a lack of understanding of the statistical methods used and of scientific procedure in general. To quote from Kallmann (p. 15): "However, this circumstance did not lead to their being reinstated in any of the principal categories after they had once been dropped for clinical reasons."

8. As regards the statistical points introduced irrelevantly into this section

by Pastore, the criticisms that eight of the secondaries have been classified as probands and that three secondaries were counted twice, are refuted by an understanding of the statistical techniques so lucidly set forth in Kallmann's book.

9. From my own prolonged observation of Kallmann's methods in the course of trips with him in New York State, I can vouch for the fact that his investigation of spouses and other relatives of probands is as thorough, and that his diagnostic procedure is as rigorous, as in the case of probands themselves—resulting in complete diagnostic uniformity.

Sampling

1. Perusal of *The Genetics of Schizophrenia* and a study of the tables will convince the reader that Kallmann includes in his sample the relevant sample characteristics, such as the distribution according to the form of schizophrenia, which his critic alleges he omits.

2. The belief that aged schizophrenics who have been institutionalized for decades have remissions shows Pastore's ignorance of the clinical aspect of the entity under consideration. Any effect that such a factor could have is thus precisely nil. Even if such a factor existed, the reason for Pastore's regarding it as a selective factor is obscure.

3. Kallmann excluded only cases in which exogenous factors played a major part, where the factor was either of a physical nature (e.g., alcohol) and the psychosis was clearly best classified in terms of the physical etiological agent (e.g., alcoholic psychosis), or where he had to do with a clear psychoneurosis.

4. Kallmann's scientific honesty in dropping forty cases because of "doubts regarding the unassailable certainty of the original diagnosis of schizophrenia" is again turned against him by his critic. An answer to the criticism appears in the text of Kallmann's book: "The investigations of their heredity and fertility followed the same lines as in the other proband groups, and even produced series of secondary cases" (p. 15).

5. That children can only originate from fertile parents is an axiom, not a selective factor. A thorough investigation into illegitimate fertility was made, and a comparison with absolute and legitimate fertility was drawn (pp. 51–66). To have sought to study "taint" in ancestry and "taint" in siblings of unmarried probands would have thrown no additional light on the nature-nurture problem in schizophrenia, and the conclusions for illegitimate children would obviously have been less reliable—smaller numbers and less certainty of parentage—than for the legitimate children.

6. The reasons for the high death rate of the children of probands below the age of 20 (chiefly due to high death rate from tuberculosis—five times that of the general population in the second decade of life) *are* considered by Kallmann, and the statistical consequences carefully evaluated and allowed for.

7. Two thousand one hundred and twenty proband children, including 111 secondary cases, are surely a sufficient number for statistical purposes, as a fourfold subdivision (into the four standard subgroups of schizophrenia) is the maximal subdivision undertaken by Kallmann in his statistical analysis. On page 21 Kallmann explicitly recognizes the inadvisability of dividing the material into too many subgroups, to avoid numbers too low for statistical reliability.

Tabular and Statistical Presentation of Data[3]

1. Enough has already been said to indicate how well versed Kallmann is in statistical method, notably the specialized statistical techniques applicable to human genetic problems.

2. Although I have studied Kallmann's tables (including Table 10) with extreme thoroughness, I am unable to agree that there are any incorrect titles or percentages with a wrong base.

3. Pastore's drawing attention to the discrepant distributions of the number of schizophrenic children, according to the form of schizophrenia in the proband, derived from the data in Tables 34 to 37 and those in chapter V, rests on his having fellen into the grave error, which he perpetrates so frequently, of failing to discriminate between net and expectancy figures.

4. The numbers included in the subgroups are invariably given, in addition to percentages.

5. Where, in certain tables, cell entries are small, they are never used for establishing statistically significant similarities or differences.

6. Most careful study has failed to reveal to me any instances of the alleged "overlapping categories." They are a fallacious postulate of Pastore, as will be explained later.

7. Two cases, and not one as alleged by Pastore, are characterized as incipient by Kallmann (case 60, p. 182, and case 89, p. 191). Perusal of the clinical descriptions of these two cases provided by Kallmann indicates clearly that although not advanced cases they are certainly definite schizophrenics.

8. The criticism that in selecting final expectancy figures, where differential choice arose, Kallmann consistently chose the higher figure and that he rejected the more accurate method on the grounds that it was more complex is simply untrue. Kallmann's criticism of Ilse's method is on purely logical and mathematical grounds, which he sets out, including Strömgren's arguments for finding Ilse's technique unsatisfactory (p. 142).

9. Kallmann sets forth clearly the expectancy figures for the three main statistical techniques employed, both in the body of chapter IV and in his conclusions (conclusion 6, p. 163). Similarly, the reader is given, at first hand, figures regarding "definites" and "doubtfuls" on which to base an opinion. For his own decision on these points Kallmann gives clear logical and mathematical reasons.

10. Pastore's claim from his own discrepant distribution as to the subgroups of schizophrenia from that given by Kallmann, and his allegation that there is a discrepancy between the biographical and tabular information on the incidence of children of female simple schizophrenics, are based on his failure to distinguish between net and expectancy figures. Pastore generalizes from the alleged error in the expectancy table for children of female simple schizophrenics, to the unreliability of other of Kallmann's tables, when the error is really his!

Evaluation of Kallmann's Major Conclusions[4]

Although certain of Kallmann's major conclusions are included in his critic's list, the principle of choice appears to have been that they are those the critic thought he had refuted, rather than those resulting from systematic selection of the major conclusions of the work.

---

[3] The numbering in this section does not correspond to Pastore's.

[4] The numbering in this section corresponds to Pastore's, with an additional point 7 in reply to his peroration.

1. Kallmann makes it quite clear, in the instance alluded to, that the 10.4 figure refers to the number of probands who had a schizophrenic parent and not to the incidence of schizophrenia among the ancestors. On p. 164, in conclusion 14, he writes: "No more than 10.4 per cent of the proband-parents could be ascertained to have been definitely schizophrenic."

2. In the matter of schizophrenic children in the *S* parental subgroup, Pastore has once again erred through confounding expectancy with net figures.

3. Nuclear and peripheral groups of schizophrenia. (*a*) In view of Pastore's error in 2, his criticism, based on this error, that the distinction between the nuclear and peripheral categories becomes confused, falls away. (*b*) Once again in this section Pastore confuses net and expectancy figures. (*c*) The attempt to explain the lower numbers of proband children in the peripheral as compared with the nuclear group in the environmentalistic terms of the lesser impairment of "family structure" resulting from the later breakdown of the schizophrenic parent, is merely a conjecture without any factual verification. In contrast to this, Kallmann's biological interpretation is based on detailed empirical evidence. Moreover, it finds confirmation by interlocking with other exact findings in the fields of genetics and constitution in schizophrenia. A further point is that Kallmann's study as a whole proves that environmentalistic factors of the type here mentioned cannot cause schizophrenia, either of the nuclear or of the peripheral variety.

4. The statistical and diagnostic points referred to by Pastore in treating of final expectancy figures in various groups of blood relationship have already been dealt with. A further misstatement introduced by Pastore is that the figure of 16.4 per cent for proband children is inflated by the inclusion of "doubtful" cases. That this is not so can be confirmed by the reader for himself by studying Table 38 in conjunction with Tables 34 to 37. Each of the latter gives the probability of schizophrenia and schizoidia in the children of probands of one of the four clinical subgroups of schizophrenia. In these tables columns for "definite" and "doubtful" schizophrenia appear side by side. In Table 38 all four subgroups are combined but only the figures for the definites are taken, and the total of 16.4 per cent is arrived at from them.

5. (*a*) Pastore has failed to grasp that the percentages for schizophrenia and schizoidia in the children of two schizophrenic parents are expectancy and not net figures. Once this is grasped, the criticism of the data yielding a figure of over 100 per cent falls away, as does the allegation of overlapping categories. (*b*) There is no excuse for Pastore's doubt concerning the group to which the figure of 9.1 per cent refers, as Table 60 on page 168 is clearly headed "Probability of Schizophrenia in Proband-Siblings." It can certainly not be said of this group that "errors due to partial accumulation of data are probably the largest."

6. (*a*) The only precise criticisms brought against Kallmann's claim of a gene-coupling between the hereditary predispositions to schizophrenia and to tuberculosis are that the schizophrenic figures included the "doubtful cases" and that the selection of the 16.4 per cent expectancy figure was made on the inadmissible grounds of choice because of the lesser complexity of the less accurate statistical method. We have already refuted both these contentions in different contexts.

(*b*) The allegation that Kallmann did not set forth the reasoning that led him to infer that there is a genuine gene-coupling is simply untrue. On page 246 there is a table (76) summarizing the salient numerical data, and on page 247 he sets forth his argument as follows:

> 1. the mortality rate from tuberculosis is highest in the descent group having the highest expectancy for schizophrenia, namely, among the proband-children;
> 2. the frequency rates for both disease groups in the other categories of blood-relationship decrease gradually;
> 3. these declining curves take an almost completely parallel course.
>
> The correspondence between the two series of figures is so far-reaching that it entirely precludes coincidence, and even exceeds the possibility of purely numerical correlation. How extensive this agreement actually is, may be demonstrated beyond question by the calculation of the proportion between the respective frequency figures. We then see that:
>
> 4. the ratio between the figures for proband-siblings and proband-children is practically identical, both in expectancy of schizophrenia and mortality from tuberculosis, and comes to 0.7 per cent in both cases;
> 5. even the corresponding ratio between grandchildren and children of probands ranges only from 0.26 per cent to 0.36 per cent.
>
> This mathematical result forges the last and most important link in the chain of our systematic taint-study in the descent of our probands. It allows hardly any other interpretation than the assumption of the *closest biological relationship between the predisposition to schizophrenia and tuberculosis*, AND above all refutes the possibility of an explanation on the grounds of similar manifestation conditions in both disease groups. At the very least, it indicates a definite gene-coupling between the tendencies to these two diseases, points to an identical pattern of heredity, and confirms the conclusion, partly established by our previous results, that both schizophrenia and tuberculosis represent *recessive* genetic factors.

(*c*) The calculation of the expectancy figures for schizophrenia and for mortality from tuberculosis *was* based on the same group of persons. It shows, moreover, a lack of understanding of scientific and statistical method to suppose that the same group of persons is necessary to demonstrate gene-coupling: witness the case of mapping the chromosomes of Drosophila where a number of different samples were employed.

7. (*a*) It is pathetic to read Pastore's statement that "a striking result in Kallmann's investigation (not brought out by Kallmann) is the finding that the offspring of probands developed only schizophrenia and no other psychosis," when a whole chapter of twenty-four closely printed pages headed "Frequency of Psychopathologic Traits Other than Schizophrenia and Schizoidia in the Descendants of the Probands" appears in Kallmann's book. Kallmann's central conclusion from the considerations adduced in this chapter is that "there are absolutely no biological or hereditary relations between the heredity circle of schizophrenia and the other abnormalities."

(*b*) It is my duty to draw attention to the crudity and inaccuracy of the formulation "the offspring of probands developed only schizophrenia and no other psychosis." The point at stake is whether the incidence of other mental disorders (making due allowance for selective factors) is *any higher* in the offspring of schizophrenics than in the general population.

(*c*) Pastore gathers that the offspring of probands develop, on the average, the same form of schizophrenia as the parents. This statement further increases one's growing conviction that Pastore has not read *The Genetics of Schizophrenia* with understanding. Kallmann's actual conclusion is "that only about half of the schizophrenia in children and grandchildren of our probands correspond to the disease form of the related probands." The evidence upon which this is based is summarized in Table 56, and the argument against different hereditary predispositions for the various schizophrenic subgroups is found on pages 149 and 150.

(*d*) Pastore states that there are sex differences for such hereditary predispositions. He apparently lacks the genetic knowledge to specify whether he is postulating sex-linkage or sex-limitation. Suffice it to say that Kallmann, after adducing a vast array of numerical data, concludes:

> The percentages for schizophrenia and schizoidia in all groups of our proband-children, reveal such thoroughgoing *independence of the sex* of the parental proband that it seems unnecessary to subdivide the other descent groups of our survey similarly, according to the sex of the proband (p. 109).

A more detailed statistical analysis of this point is to be found on page 124.

Summary on Pastore's Review

Even had Pastore proven all his points instead of making a series of *faux pas* through ignorance of genetic and statistical methodology, his sweeping generalization (whereby he sets at nought a thorough, systematic, genetically and psychiatrically enlightened study of such proportions that it took ten years to complete) would have been untenable: the hereditary nature of schizophrenia and even the type of genetic mechanism deduced by Kallmann would not have been upset, for Pastore emphasizes minutiae, neglecting the general effect of the vast body of Kallmann's data which he does not criticize.

I trust that this reply of mine has succeeded in showing up the fallaciousness even of Pastore's criticism of the minutiae, resulting from his lack of training in the disciplines (genetic ideology and population statistics) relevant to the type of study under consideration.

Confirmatory Evidence for Kallmann's Viewpoint

In conclusion, let me adduce confirmatory evidence for Kallmann's salient findings and viewpoint by (1) referring to his own more recent studies (twin-family) in schizophrenia and tuberculosis, and (2) making a general summary statement of conclusions from studies of other workers in this field.

1. *Kallmann's later work on schizophrenia and tuberculosis.* (*a*) *Schizophrenia.* In November, 1946, Kallmann published his analysis of 691 American schizophrenic twin index families (6), which confirmed

in all ways the salient findings of his German study, on which *The Genetics of Schizophrenia* was based. One of the conclusions of this paper—namely, that the genetic theory of schizophrenia "is equally compatible with the psychiatric concept that schizophrenia can be prevented as well as cured"—may appear paradoxical to those unacquainted with the concepts of modern physiological genetics. So far from equating the notions of heredity and irreversibility, this science shows how by the identification of the organic, biochemical substratum of hereditary and constitutional anomalies, a direction is given to research, which now has a tangible, more limited field upon which to concentrate. Replacing the hereditarily deficient chemical in diabetes mellitus is a case in point. Viewed in this light it will be understood how heredito-constitutional abnormalities may be regarded as holding out a better prospect of ultimate cure than psychogenic troubles, which are at the mercy of an environment which for the individual can never wholly be controlled. The heredito-constitutional view of schizophrenia has already borne fruit in the realm of practical application to therapy. Not only does it offer a rationale for the efficacy of shock treatment in terms of strengthening the functional efficiency of the reticulo-endothelial system, which is thereby better able to protect the brain against noxae (possibly endocrine in nature), but it also explains the effects of exercise in the open air and of weight regulation on the progress or arrest of schizophrenia. Kallmann has developed this theme in certain recent papers (**7, 8, 9**).

(*b*) *Tuberculosis.* Two articles (**10, 11**) written by Kallmann in collaboration with David Reisner, head of the Bureau of Tuberculosis, New York City Department of Health, confirm Kallmann's earlier conclusions as to a specific hereditary predisposition to tuberculosis, and elaborate the particulars of a multifactorial genetic mechanism modifying the resistance to this disease. The material investigated comprised 657 twin pairs and their families, from hospitals and clinics in the State and City of New York, reported over a period of about five years.

2. *The findings of other investigators.* The following is a summary statement of the findings of other investigators confirming the correctness of Kallmann's genetic view of schizophrenia: (*a*) No population survey has hitherto yielded a general expectancy rate for schizophrenia of over 1 per cent; (*b*) no study of a representative group of blood relations of schizophrenics has failed to yield a *significant increase* in this expectancy rate.

BIBLIOGRAPHY

1. Hurst, L. A. Seventh International Congress of Genetics. *Ment. Hyg., N.Y.*, 1939, **23**, 677–682.
2. Hurst, L. A. Heredito-constitutional research in psychiatry. *South African med. J.*, **14**, 1940, 384–400.
3. Hurst, L. A. Hereditary factors in mental disorder and mental defect. *South African med. J.*, 1944, **18**, 397–401.
4. Kallmann, F. J. *The genetics of schizophrenia*. New York: Augustin, 1938.
5. Kallmann, F. J. The scientific goal in the prevention of hereditary mental disease and race inferiority. *Proc. Seventh Int. Congr. Genet.*, Edinburgh, Aug. 23–30, 1939.
6. Kallmann, F. J. The genetic theory of schizophrenia. An analysis of 691 schizophrenic twin index families. *Amer. J. Psychiat.*, 1946, **103**, 309–322.
7. Kallmann, F. J. Modern concepts of genetics in relation to mental health and abnormal personality development. *Psychiat. Quart.*, 1947, **21**, 535–553.
8. Kallmann, F. J. Genetics in relation to mental disorders. *J. ment. Sci.*, 1948, **94**, 250–257.
9. Kallmann, F. J. Applicability of modern genetic concepts in the management of schizophrenia. *J. Hered.*, 1948, **39**, 339–344.
10. Kallmann, F. J., & Reisner, D. Twin studies on the significance of genetic factors in tuberculosis. *Amer. Rev. Tuberculosis*, 1943, **47**, 549–574.
11. Kallmann, F. J., & Reisner, D. Twin studies on genetic variations in resistance to tuberculosis. *J. Hered.*, 1943, **34**, 269–276.

Received October 12, 1950.

CORRELATION VERSUS CURVE FITTING IN RESEARCH ON ACCIDENT PRONENESS: REPLY TO MARITZ

MILTON L. BLUM AND ALEXANDER MINTZ
City College of New York

Maritz (**3**) states that the technique of correlating accident records in two successive periods is indispensable as evidence of accident proneness. He suggests that the fitting of theoretical Poisson and Negative Binomial distributions is not an adequate criterion of the absence or presence of differences in accident proneness in a group. There are certain weaknesses in this position, and a clarification is demanded. To reiterate the major points of our earlier paper (**4**):

1. Personal accident proneness, a component of accident liability, has been overemphasized.
2. This can be demonstrated by a method which reveals the extent to which accident records *could* be attributed to differences in accident liability, and this was found to be 20 per cent to 40 per cent of the total variance of accident records.
3. This method is based on the use of univariate distributions.

We agree with Maritz that a good Negative Binomial fit does not prove the existence of differences in accident proneness with mathematical certainty. We never said that it did. It is doubtful whether anything is ever proved with mathematical certainty in empirical sciences.

The issue raised by Maritz is that the correlation technique is indispensable in the establishment of accident proneness and that the evidence from a univariate distribution is invalid. We strongly disagree. We shall show that the two techniques give much the same information and, therefore, either can be offered as evidence for accident proneness. In fact, neither is wholly conclusive. Detailed histories of individuals' accident careers should be better than either.

Maritz views correlations as the "direct technique" of establishing differences in accident proneness. His claim is that "the most direct method of establishing proneness in a group of people all of whom ought to be exposed to the same environmental risk, consists of splitting a lengthy period of observation into two periods and correlating the frequency of accidents per individual for these two periods. This statistical technique is nearest to the psychological definition of accident proneness." He attempts to show that this technique may contradict conclusions based on univariate distributions. His evidence is based upon two hypothetical distributions and a previously unpublished distribution by Adelstein. One of his hypothetical distributions resembles a chance pattern and yet yields a correlation in successive periods. The other is suggestive of differences in accident proneness but is uncorrelated in successive periods. Adelstein's data, in his opinion, illustrate two simple chance distributions which are correlated with each other.

Maritz' two hypothetical distributions illustrate the mathematical possibility for two Poisson distributions to be correlated, and for a Negative Binomial Distribution to result from summation of two uncorrelated distributions. However, the occurrence of both kinds of bivariate distributions in the case of accident records is unlikely. The mathematical derivation of the bivariate correlated Poisson distribution which Maritz applies to accident records shows only that it can be approximated by drawing colored balls from enclosed boxes. We do not believe that such a distribution should be expected in the case of accident records. The only obvious meaning of a Poisson distribution of accident records is that it is satisfactorily explained by the assumption of equal and constant accident liability. Without such an assumption it looks like a result of an odd coincidence. If Poisson distributions of accidents are the results of equal liability, they should be uncorrelated, and it is not clear what kind of combination of circumstances should lead to the expectation of Poisson distributions correlated in successive periods.

In his other hypothetical distribution this lack of plausibility is obvious. For example, all six hypothetical individuals who had more than eighteen accidents each in one period had zero accidents in the other period. This could occur, but is hardly to be expected. In other words, in this example, Maritz proves that it is possible to obtain a correlation of −.11 by arranging numbers in a manner designed to obtain it. This is granted, but what does it prove about accidents?

The only empirical material Maritz presents is the unpublished Adelstein data, and it must be regarded with more seriousness than his two hypothetical distributions, which are mathematically possible but highly improbable. Maritz claims that the examination of the univariate distributions suggests a "pure chance pattern," but that the correlation between the accident records in the two periods reveals the existence of differences in accident proneness to the extent of a correlation of .29. The existence of the correlation is treated as something that could not have been foreseen in terms of the accident distributions in the two observation periods. Thus, the impression is created that the combined period had properties which were essentially different from those of the shorter periods when considered alone. He bases his interpretation on the fact that three χ^2 tests fail to reveal significant differences either between two Poisson distributions and Adelstein's five- and six-year distributions, or between a bivariate correlated Poisson distribution and Adelstein's scatter diagram for the two periods.

In dealing with these data Maritz makes the common error of confusing the failure to disprove a hypothesis with its proof. He states: "Equation [1] was fitted to the observed data of Table III and the resulting test of goodness of fit gave for 7 *df*, $P = .49$. Hence it follows that the above data follow a *correlated* bivariate Poisson distribution" (p. 438). This second sentence does not follow from the first. The failure to disprove a hypothesis according to which the data for a combined period have properties (the .29 correlation) different from those of the constituent periods (viewed as exhibiting a simple chance pattern) is not the same as proof of it. A closer examination of the data for direct evidence of this type of heterogeneity fails to reveal anything convincing, and the opposite and theoretically more plausible hypothesis of essential homogeneity of the eleven-year observation period fits the data more closely than Maritz' heterogeneity (i.e., bivariate correlated Poisson) hypothesis.

In terms of the properties of the accident distributions in the two consecutive periods, the most probable correlation between the accident records in these periods is not zero as Maritz implies, but .21. This is quite close to the observed .29. The estimate of a correlation of .21 was arrived at by determining the estimated percentages of the accident records attributable to factors other than chance (18 per cent and 24.4 per cent) and then computing their geometric mean. In determining these percentages the formula $(v-m)/v$ was used; the two variances were 1.485 and 1.370, respectively, and the two means were both 1.123.

If the eleven-year period had no properties essentially different from those of the shorter period, it should be possible to construct from theoretical scatter diagrams approximating the empirical one by utilizing statistics derived either from the two periods considered separately (without considering their correlation), or from either one of them taken alone, or from the total period taken as a whole. The form of the bivariate distribution chosen was the bivariate Negative Binomial. It is based on much the same assumptions as those made by Greenwood and Yule (1) in their derivation of a univariate unequal liability distribution, namely:

> Accident liability is distributed in people in accordance with a Pearson III curve.

Accident liability of a person remains constant per unit of time throughout the two observation periods.

Each particular degree of accident liability gives a simple chance (Poisson) distribution of accident records in each observation period (p. 279).

Table I presents the Adelstein data together with a theoretical distribution constructed in accordance with these assumptions; the computation utilizes only the mean and the variance of the first period, and the fact that the second period lasted six years while the first one lasted five years. The formula for the theoretical frequency for the cell representing j accidents in the first, k accidents in the second is

$$y_{jk} = Na^k\left(\frac{c}{c+a+1}\right)^p \frac{\Gamma(p+j+k)}{\Gamma(p)j!k!(c+a+1)^{j+k}}$$

in which a is the ratio of the two durations, and c and p are two constants derived from the mean and variance of the first period, as follows: $c=m/(v-m)$, $p=m^2/(v-m)$. (The derivation of the formula, which closely resembles that of Greenwood and Yule, will be published elsewhere.)

TABLE I

A COMPARISON OF ADELSTEIN'S ACTUAL ACCIDENT DATA WITH THE THEORETICAL BIVARIATE DISTRIBUTION COMPUTED FROM HIS FIRST-PERIOD DATA ($m=1.123$, $v=1.370$)

First Period	*Second Period*								
	0	1	2	3	4	5	6	7	Total
0	21(16.3)	14(14.8)	8(8.0)	1(3.3)	—(1.2)	—(0.4)	—(0.1)		44(44.1)
1	17(12.3)	12(13.4)	8(8.4)	3(4.1)	1(1.6)	—(0.6)	—(0.2)	1(0.1)	42(40.7)
2	6(5.6)	9(7.0)	2(5.1)	2(2.7)	2(1.2)	—(0.5)	—(0.2)	—(0.1)	21(22.4)
3	1(1.9)	1(2.8)	3(2.3)	3(1.4)	1(0.7)	—(0.3)	—(0.1)		9(9.5)
4	1(0.6)	3(1.0)	—(0.9)	—(0.6)	—(0.3)	—(0.1)			4(3.5)
5	—(0.2)	—(0.3)	—(0.3)	2(0.2)	—(0.1)				2(1.1)
6	—(0.1)	—(0.1)	—(0.1)	—(0.1)					0(0.4)
Totals	46(37.0)	39(39.4)	21(25.1)	11(12.4)	4(5.1)	0(1.9)	0(0.6)	1(0.2)	122(121.7)

The theoretical distribution appears to fit the data quite well. The χ^2 computed was 8.743, $df=10$, $P=.58$. It follows that the correlation technique recommended by Maritz did not add anything significantly new to the information one could gather by examining one of Adelstein's univariate distributions.

In arguing for correlations and against properties of univariate distributions as grounds for assuming differences in accident proneness, Maritz overlooks the fact that these two kinds of statistical measures are closely interrelated. The correlation between accident records in two periods can always be computed from the variances in these two periods and in the total period, according to the formula

$$r = \frac{V_{1+2} - V_1 - V_2}{2\sqrt{V_1 V_2}}.$$

Similarly, the increase in accident variance when one combines two observation periods is an increasing function of the correlation between these periods, in accordance with the elementary formula $V_{1+2} = V_1 + V_2 + 2r\sqrt{V_1 V_2}$. Inasmuch as this formula is exact, and inasmuch as every observation period has an early stage when the variance is smaller than the mean, a distribution cannot have a variance greater than the mean value unless there were positive correlations between successive periods somewhere in the past. Similarly, a Poisson distribution cannot result unless the accident records in the subdivisions of the observation period were uncorrelated or unless the effects of positive and negative correlations cancelled each other. If an examination of two univariate distributions in two successive periods suggests something markedly different about the existence of accident proneness, compared to the correlation between these periods, the interperiod correlations must have similarly changed in the past. Maritz' hypothetical distributions could be used just as readily in arguing against the use of correlations in accident research as against the use of variances. The only advantage of a correlation lies in the fact that it enables one to tell a particular time when the variance of a set of accident records has risen at a rate beyond chance expectation.

A further consideration with reference to the correlational technique is that it presents certain practical difficulties. Accident proneness is a problem for industry as well as for the theoretical statistician. From the point of view of industry it is often difficult to acquire data on the same individual for two successive periods. Those with high accident rates in the first period are likely not to be found in the second period. They may be dismissed or resign from their jobs, not to mention being hospitalized or dead. The study reported by Kerr (**2**) is typical of practical problems confronting industry. The major effort is to reduce accidents, not to wait for successive periods.

SUMMARY

1. The hypothetical distributions presented by Maritz are mathematically possible and demonstrate the lack of mathematical certainty

of inference from empirical data, but such distributions are not likely to be encountered in practice.

2. Correlational research on accident proneness is legitimate, but inferences about accident proneness drawn from correlations are not more certain than inferences drawn from the fitting of distributions.

3. Correlational research is not always feasible for practical reasons. In any event, it is not indispensable with reference to establishing accident proneness.

BIBLIOGRAPHY

1. GREENWOOD, M., & YULE, G. U. An enquiry into the nature of frequency distributions representative of multiple happenings, with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *J. roy. statist. Soc.*, 1920, **83**, 255–279.
2. KERR, W. A. Accident proneness of factory departments. *J. appl. Psychol.*, 1950, **34**, 167–170.
3. MARITZ, J. S. On the validity of inferences drawn from the fitting of Poisson and Negative Binomial distributions to observed accident data. *Psychol. Bull.*, 1950, **47**, 434–443.
4. MINTZ, A., & BLUM, M. L. An examination of the accident proneness concept. *J. appl. Psychol.*, 1949, **33**, 195–211.

Received November 18, 1950.

CONCERNING TASTE-BLINDNESS TO PTC

JOZEF COHEN
University of Illinois

AND

DONALD P. OGDON
University of Texas

A recent paper by Boyd (**1**) has offered several criticisms of a review paper on taste-blindness to PTC published by Cohen and Ogdon (**2**). This note will reply to those criticisms, using the same paragraph numbering as Boyd's paper. References will not be repeated since they may be found in the bibliographies of the above communications.

1. In the first paragraph of his paper, Boyd has said that our report that Lasselle and Williams were the first discoverers of taste-blindness (to creatine) is false. Boyd indicates that he investigated this substance shortly after publication by Lasselle and Williams and found that no taste-blindness existed. The trouble with this sort of affair is, that all things being equal, we believe what is reported in the literature. We read what Lasselle and Williams had found, and we included it in our review. We did not read what Boyd had found for the reason *that he did not put it into the literature.* Since we didn't read it, we didn't mention it.

2. In this section Boyd refers to a "careful study" by Hartmann as finding that non-tasters could not taste PTC even when it was dissolved in saliva from strong tasters. This study, says Boyd, should have been included in our review.

Miss Hartmann did not come to the above conclusion. She found, in fact, that one non-taster (herself) could not taste PTC when it was dissolved in the saliva of one taster. The *entire* study is reported in a single sentence; there is not the slightest mention of concentration, technique, nor is there the slightest mention of a control. The total N was equal to 1. We consider the description of the experiment as a "careful study" as being ludicrous, and reject the suggestion that it should have been included in our section concerning the role of saliva in the tasting of PTC.

3. Here Boyd discusses sex differences in taste-blindness to PTC. Boyd says that we should have included three studies in support of his claim that sex differences are significant. These studies are Boyd and Boyd, Falconer, and Riddell and Wybar. We specifically cited the results of Boyd and Boyd in our paper. We had not seen the Falconer

paper at the time we wrote the review. The reason is that the review was written in 1948, and the Falconer paper was published in 1947 in England. It takes about a year for material printed in England to be brought to America and abstracted by the appropriate agencies. It had not come to our attention.

The suggestion that we should have included the Riddell and Wybar study in support of sex differences in taste blindness is something we call to the attention of the careful scientist. Riddell and Wybar state clearly and precisely that their data *show no significant differences.*

4. Boyd discusses here the racial distribution of tasters with reference to our table containing, by far, the majority of studies completed in this area. We maintain two things: (*a*) the numbers do not form racial clusters, (*b*) the table is inconsistent. Racial clusters are not obvious, and we have yet to see a map drawn with accurate contour lines. The table is inconsistent because American Caucasians have values all the way from 60 per cent to 82 per cent. We did not say that there are not significant differences between groups indicated in the table. Significant differences do occur, probably even between different experiments on what are supposed to be the same group—American Caucasians.

We are glad to learn that Boyd has developed a standard technique for testing for taste to PTC. His paper describing the more important aspects of it was published *after* our review appeared; we were unable to include the technique because we were writing about PTC and not clairvoyance. In any event, other investigators did not use Boyd's technique, and their results, as we have said, are not directly comparable.

5. This paragraph, recapitulating our reports of taste-blindness to di-phenyl-guanidine and thiouracil, is based on a 1950 paper by Boyd and an unpublished manuscript by Boyd and Hoffman, and we will not comment on it. A review paper, or a criticism of a review paper, should not contain any new material, and we object to criticisms involving experiments which were unpublished at the time of publication of the review paper.

BIBLIOGRAPHY

1. Boyd, W. C. "Taste blindness" to phenylthiocarbamide and related compounds. *Psychol. Bull.*, 1951, **48**, 71–74.
2. Cohen, J., & Ogdon, D. P. Taste-blindness to phenyl-thio-carbamide and related compounds. *Psychol. Bull.*, 1949, **46**, 490–498.

Received March 27, 1951.

REPLY TO TRAVERS' "A CRITICAL REVIEW OF THE VALIDITY AND RATIONALE OF THE FORCED-CHOICE TECHNIQUE"[1]

DONALD E. BAIER[2]
Department of the Army

The Personnel Research Section of The Adjutant General's Office is an operating research agency. It conducts both basic and applied research for the purpose of providing the Army with the best possible personnel tools. As an operating research agency, it sometimes, because of the pressure of events, delays publication of results of interest to the psychological profession. By the time research is published, all too often progress has been made toward further refinement or, sometimes, even toward a different viewpoint on a given problem. The recent review by R. M. W. Travers of the forced-choice technique (**23**) is a case in point.

Travers is in the unfortunate position of attempting a review of a problem which originated in the Personnel Research Section and on which but a small fraction of the Section's research has been published in the psychological journals. However, reports of additional research are available as Personnel Research Section Reports[3] and as papers read at meetings of the American Psychological Association. This additional information would have kept Travers from making such statements as the following (**23**):

1. The following quotation summarizes all the statistical evidence that the present writer has been able to obtain concerning the relative validity of the forced-choice (FCL) and the more traditional type of rating device (RCL) as it is used in the Army Officer Efficiency Report (p. 67).
2. The claims for the validity of the technique seem to bear little relationship to the actual evidence (p. 66).
3. Proper studies need to be made to determine the validity of scales in this area constructed on the basis of an adequate rationale (p. 70).

Travers could have obtained the additional information had he but indicated his intent to write a review. The difficulties of generalizing in the field of personnel research, resulting from the operation of so many factors, are well known. A critical review is scarcely worth the name if it leans heavily on a single study and does not cover other research information.

---

[1] TRAVERS, R. M. W. A critical review of the validity and rationale of the forced-choice technique. *Psychol. Bull.*, 1951, **48**, 62–70.

[2] The opinions expressed herein are those of the author and do not necessarily reflect the views of the Department of the Army.

[3] Personnel Research Section Reports are not available for general distribution. Arrangements have been made to furnish the American Documentation Institute with copies of unclassified reports for distribution.

There is one highly important omission from Travers' article. His title and his remarks are directed at the forced-choice technique in general, although the data he discusses are confined to rating procedures. It should not be overlooked that the forced-choice technique has been used in other types of instruments, for example, personality inventories and self-description forms. The forced-choice technique has demonstrated its great usefulness in the construction of such instruments (**12**, **13**, **16**, **19**). Even higher validities have been obtained with modifications of conventional forced-choice technique based on suppressor theory (**1**, **20**). The lack of explicitness in Travers' review on this point is misleading. Compared to the traditional "yes-no" type of questionnaire, the technique has produced personality measures of useful predictive value for the Army situation. Working with the traditional type of personality items has consistently failed to yield a useful product.

The comments in this reply refer only to rating scales used for efficiency-reporting or merit-rating purposes as involved in the work of the Personnel Research Section.[4] Further, the paper is not intended as a complete review of the research concerning this application of forced-choice. It is an effort to discuss the problems raised by Travers, and to make available some of the more general findings concerning the problem of efficiency reporting. This effort will be made in terms of: (1) areas in which there is agreement with Travers, (2) areas in which there is disagreement with him, and (3) areas or problems concerning which he makes no comment.

AREAS OF AGREEMENT

Ratings should not be validated against other ratings (**23**, p. 69). From one point of view there is no doubt of the desirability of "criteria other than judgments." The Personnel Research Section and other investigators have searched and still are searching for more objective and appropriate criteria. Practicable suggestions for the development of such criteria would be eagerly welcomed. Until such a development occurs, investigators will no doubt continue to use ratings as criteria, and considerable effort will be expended toward improving such ratings.

The problem of criteria for efficiency reports deserves more than the brief comment Travers gives it. In considering the use of ratings versus objective criteria in any instance, the nature of "success" being predicted must be carefully considered. In some instances, ratings may be the best criteria because value judgments are the essential elements. What must be clearly recognized also are the problems of interpretation that are involved when ratings are used to "validate" ratings. More specifically, three points should be noted concerning the use of a composite of ratings as a criterion.

---

[4] The problems involved in securing ratings for criterion purposes are not necessarily the same.

1. It is considerably better than no criterion at all. It is well known that averaging a series of ratings will tend to reduce bias. At the very least, therefore, use of multiple ratings as criteria in evaluating rating scales will improve rating procedures by identifying those scales which contain the least amount of bias.
2. Use of composite ratings as a criterion would seem to have its maximum justification in studies of performance as an officer. This is the case, since performance as an officer involves, as a large component, the ability to work with and through other people. Furthermore, an officer's career involves a large variety of duty assignments; the expression of his value must be in generalized terms. Judgments of superiors, subordinates, and immediate associates are especially pertinent.
3. Use of multiple ratings as criteria creates problems of interpretation of the findings involving comparison of specific rating techniques. Up to the present, the rating composite has essentially been an averaging of ratings obtained by a single technique—the traditional type of rating scale. When a rating scale is involved as a predictor, one never knows the extent to which it is favored because of its similarity to the criterion. Indirect evidence suggests that the amount of such "technique contamination" is appreciable (**17**). The solution to the problem of comparing rating techniques, when criteria differing entirely in character are unavailable, may be the inclusion of all types of rating techniques in the criteria. This procedure will give each rating technique an equal opportunity of showing "validity." Such a procedure reduces the problem of "validity" in rating studies to one of rater agreement, i.e., reliability, if this concept is considered to cover a relationship of one rater using a given technique to several raters using the same technique.

In the sense of choosing between two members of a pair, forcing a choice is not an essential part of the technique. In discussing the rationale of the technique, Travers makes a great deal of the point that all items of a pair or a tetrad could either be listed in rank order or the rating could be given in terms of a traditional rating scale with the restriction that no two traits could be rated at the same point (**23**, pp. 64–65). This may well be true. The possibility has already been indicated in conjunction with self-rating items (**7**, p. 186). Inasmuch as it is obviously difficult, if not impossible, to extend indefinitely the number of items which can be considered together, an element of forcing is bound to be present. This is nothing new. Choices must be made among the terms that are grouped together, much in the same sense that a choice must be made among the alternatives of any multiple-choice item. This point is relatively unimportant except for its relationship to the next.

Forced-choice pairs work because the nonscored alternative serves as a suppressor. This is an important point because of its theoretical significance. We agree that suppressor theory may provide the rationale for the success of forced-choice items; in fact, we have exploited it heavily in connection with self-description inventories, as mentioned above (**20**). To avoid any misunderstanding, however, certain points should be made explicit:

1. Travers states that forced-choice procedure assumes "for any given individual, the true rating on the irrelevant [i.e., unscored] items is average and that any deviation in the average ratings on these irrelevant characteristics represents a tendency to over-rate or under-rate the particular individual who is being rated . . . . The procedure is simply that of using ratings on certain characteristics as a suppressor variable to correct for errors in the rating on certain other characteristics" (**23**, p. 65).[5]

2. The above assumption is not necessary, nor was it made in the development of the forced-choice procedure. To quote from one of the Section's early papers (**11**) on this technique, "The essence of the forced-choice technique, as we use the term, however, is the grouping of the alternatives to make them appear of equal value, and yet have unequal significance." In other words, items are paired so as to give each alternative equal face validity and differing true validity. Whether or not individuals have an average rating on "irrelevant" items is not essential for either of these conditions to obtain, nor has it any bearing on suppressor theory in forced-choice items.

3. The suppressor theory requires only that (*a*) the scored alternative of a pair have as high a validity as possible; and (*b*) the nonscored alternatives have as low (even negative) a validity as is consistent with a high relationship with the scored item.

4. A casual reader of Travers' article might conclude that a separate suppressor key could be developed for the traditional rating scale items. Travers did not suggest this possibility, but it is desirable to make explicit the point that this application of suppressor theory may not work. Use of traditional ratings as suppressors for other traditional-type ratings has been tried in obtaining rating criteria. The suppressor theory has been confirmed in the sense that negative Beta weights were obtained for the intended suppressor ratings, but the effect was so slight that validity of the combined ratings was not improved (**18**). It is not intended to assert that a suppressor key might not be developed on the basis of traditional-type items, but only that available evidence does not encourage optimism in this belief.

Forced-choice items can be improved in their content (**23**, pp. 63–64).

---

[5] "Irrelevant" and "relevant" are perhaps not sufficiently meaningful in this context. In one sense, all items are "relevant" to officer performance. "Discriminative" or "nondiscriminative," "differentiative" or "nondifferentiative," "critical" or "noncritical" are suggested substitutes.

Pairing items on a statistical basis only will frequently bring together alternatives which normally would not be associated. Indeed, this may be one of the advantages of the forced-choice procedure. Travers' point that the content of the pairs should not confuse the rater is well taken. However, we believe he has exaggerated the problem. Directions to the rater have always stressed that he is to indicate which alternative *most nearly* applies to the person he is rating. Items such as Travers cites do meet the crucial test of having and *maintaining validity* over a period of time (**8**).

Forced-choice items do not prevent the rater from manipulating his rating if he so desires (**23**, pp. 69–70). In publications of the Personnel Research Section, claims have been much more modest than Travers implies. To cite one instance, " . . . it reduces the rater's ability to produce any desired outcome of obviously good or obviously bad traits. It, thus, diminishes the effect of favoritism and personal bias" (**10**). The emphasis is on the words "reduces" and "diminishes."

Personal bias is a general term indicating departure from the true value for any reason. Bias may result from insufficient information on which to base a rating, from the unconscious operation of friendship, from differences in leniency on the part of raters, etc. It is in the reduction of these types of bias that the forced-choice technique may be particularly helpful. The rater who deliberately desires to manipulate his rating can undoubtedly do so. However, the forced-choice technique makes it a somewhat more difficult task for him.

In passing, it might be pointed out that an efficiency report is primarily a means of recording the rater's estimates. By itself, regardless of technique used, it does not guarantee that the rater will be honest, comprehensive, careful, and objective. To achieve this purpose, supplementary aids must be used, and even these may not be effective. In the Army, this aid is in the form of an Army Regulation which contains not only the necessary administrative procedures but also a discussion of the purpose and use of the efficiency report and of the psychological principles involved in rating. This psychological information would not have been included in the Regulation if it were believed that the forced-choice technique were an automatic and complete control of rater bias.

In relation to this question, it should be pointed out that while a rater can move his rating up or down the traditional-type rating scale at will, and can influence the score he is giving on forced-choice items, on neither type of rating scale can he determine with much precision the relative standing of the person he is rating. This point is most clearly seen when scores on rating scales are translated into some standard scale. It is not uncommon on a seven- or eight-point rating scale for 30 per cent of the responses to be concentrated at a single point. The amount of change a swing of one point on a scale will produce on a standard score is evident. Unless, therefore, a rater knows precisely the distribution of ratings, he can never know where he is placing a person on a relative population scale, the kind of rating used by the Army. This point is mentioned because it is believed that a good deal of the objection encountered by the forced-choice technique has been misdirected, and the point to which objection is taken is basically the difficulty of reconciling relative and absolute standards.

AREAS OF DISAGREEMENT

Some of the areas of disagreement are quite minor. These will be disposed of first.

1. *Travers is incorrect in his statement, "Each one of these elements describes, in essence, a rather specific item of behavior"* (**23**, p. 62). A glance at the alternatives shows very many general terms, i.e., *modest, no one ever doubts his ability, low efficiency, businesslike.* One of the unsolved problems is the degree of specificity which alternatives in forced-choice groupings should possess.

2. *Travers is incorrect in his interpretation of preference index.* He states, "The other value [preference index] indicates the extent to which individuals tend to rate others too high or too low on a particular characteristic" (**23**, p. 62). The preference index is, to quote from an early publication on this technique (**11**), an index of the "value to the rater" of the alternative under consideration; more recently, the preference index has been considered as a measure of the face validity of the item. It is hoped that use of the forced-choice technique will tend to correct for raters' tendency to rate too high or too low, but this is not involved in the computation of the preference index.

We disagree with Travers' statement, "Claims for the validity of the technique seem to bear little relationship to the actual evidence" (**23**, p. 66). In support of this statement, Travers relies heavily on a minor study of the Personnel Research Section (**14**) and one by Richardson (**6**). As mentioned earlier, in making this statement he has ignored the vast body of research data available. Some of the data have been presented at recent meetings of the American Psychological Association (**2**, **4**, **8**, **9**, **21**).

It would take us too far afield to review this work here. Perhaps the information in Table I (from **15**), based on two samples totaling 7,771 cases, will suffice to indicate the type of evidence available to support the statement that " . . . it [a combination of forced-choice and graphic rating scales embodied in an official efficiency report] produces ratings which are more valid indices of real worth" (**10**). This table reports

TABLE I

Comparative Validity of Form 67 and Form 67-1, April, 1946, Edition (from 15)

Rank	*Sample 1* (*N*=4,208)		*Sample 2* (*N*=3,563)	
	Form 67	Form 67-1	Form 67	Form 67-1
Col.	.24	.35	.30	.30
Lt. Col.	.13	.23	.48	.50
Maj.	.32	.42	.32	.34
Capt.	.21	.31	.34	.35
1st Lt.	.34	.46	.45	.51
2nd Lt.	.30	.45	.46	.57

some of the results of a study conducted in connection with the regular reporting period of 30 June 1946. Both WD AGO Form 67 and WD AGO Form 67-1 were completed for the same officers. The score on Form 67 was an average of ten eight-point graphic rating scales; the score on Form 67-1 was a combination of forced-choice and rating scales. The criterion used was an average of ratings by superiors, subordinates, and associates obtained by a nominating technique. The consistently greater validity of Form 67-1 is evident.

We disagree with Travers' interpretation of the quotation "[the forced-choice rating technique is] relatively free from the usual pile-up at the top of the scale" (**23**, p. 66). In the first place, Travers has confused the forced-choice technique per se with Form 67-1. This form contains both forced-choice and traditional-type rating scales. The distributions he reproduces (from **10**) are for the total score on Form 67-1. Travers does not observe this distinction; hence, his remarks are misdirected.

In the second place, Travers does not comment on the difficulties in comparing a distribution based on a scale of 220 used points (Form 67-1) with a distribution based on a scale of 43 used points (Form 67).[6] The attempt to equate the range for the purpose of comparing distributions on the two forms gives the traditional rating scale (Form 67) every advantage.

In the third place, Travers has missed the point. A Personnel Research Section Report, dated 17 January 1947 (**15**) contains the information on which were based the illustrative Distributions reproduced as his Figure 1. The computation of the third and fourth moments contained in that report shows that Form 67 has greater leptokurtosis and that Form 67-1 has greater negative skew. "This will mean that Form 67-1 will be more discriminative of extreme cases than will 67, particularly at the low end of the distribution" (**15**, p. 7).

---

[6] The score on Form 67 has a possible range of −4 to +7. Scores below 2.7 are rare. Considering the score in tenths of a point gives 43 points in the actual range.

In the same Personnel Research Section Report, there appear data (**15**, p. 16) which show the percentage of officers at two cut-points on the distributions. These data are reproduced as Table II. If the equated data are taken at face value, this table clearly brings out the better discrimination by Form 67-1 at both ends of the distribution. Below the point where the curves (**23**, Fig. 1) cross at the low end of the distribution, there are 18.2 per cent for Form 67-1 and 9.5 per cent for Form 67. Above the point where the curves cross at the high end of the distribution, there are 10.7 per cent for Form 67-1 and 8.6 per cent for Form 67. Table II shows the same type of information by grade. The point of particular interest is that for the lower grades, the difference in effectiveness of the two forms is most pronounced at the high end of the scale; for the upper grades, the difference is most marked at the lower end of the scale. In the light of this kind of data, there is no question as to which form is the more useful.

TABLE II

PERCENTAGE OF OFFICERS BEYOND LOW AND HIGH CUT-POINTS ON FORM 67 AND FORM 67-1

(From **15**, Table V)

Grade	*% to Lower Cut from Bottom*			*% to Upper Cut from Top*		
	Form 67	*Form 67-1*	*% Excess*	*Form 67*	*Form 67-1*	*% Excess*
Colonels	31.46	60.28	28.82	No cross	No cross	0.00
Lt. Colonels	24.69	48.15	23.46	No cross	No cross	0.00
Majors	13.28	33.65	20.37	3.32	4.53	1.21
Captains	12.06	20.24	8.18	5.07	10.10	5.03
1st Lts.	3.56	12.93	9.37	10.32	17.22	6.90
2nd Lts.	5.51	13.53	8.02	9.72	17.89	8.17
Combined	9.48	18.16	8.68	8.55	10.67	2.12

We disagree with Travers' conclusion: "*The data [from* **14**] *suggest that experimentation with different types of directions may yield much more important results than experimentation with forced-choice scales*" (**23**, p. 68). Experience of the Personnel Research Section indicates that this statement represents an extreme oversimplification of the problem involved. In our experience, the basic attitudes of the raters, determined in large part by knowledge of the uses to which a rating scale is to be put, are little influenced by specific directions.[7] If psychological effects of such a kind influence ratings, it would appear more likely that they will be brought about by the rater's actually doing something (for example, some form of preliminary ratings) than by a certain type of instruction. Furthermore, raters do not necessarily follow the instructions. Attention is called to the fact that despite what must be an extraordinary variety of directions for merit rating forms, high negative skew and leptokurtosis are almost invariably characteristic of the ultimate distributions. These stubborn characteristics, in fact, have served as motivation for the search for other than the traditional rating techniques.

We disagree with Travers' interpretation of a rater agreement represented by a correlation of 0.69 (**23**, p. 69). While it is not the intention to discuss Travers' comments on Richardson's study, from which this "reliability" coefficient is cited, it should be pointed out that the "only 0.69" leads to the wrong evaluation of a coefficient of this magnitude. This may be illustrated from a follow-up study of the validity of Form 67-1 (**17**). In this study, rater agreement on the criterion ratings obtained at a single sitting was represented by $r=.24$. Agreement between the official Form 67-1 raters for 914 cases was represented by $r=.56$. In Army experience at least, rater agreement as represented by coefficients as high as .70 is a rare finding and not to be considered unusually low.[8]

EFFICIENCY-REPORTING PROBLEMS OMITTED FROM TRAVERS' ARTICLE

In attempting a critical review of a technique, it is customary to discuss problems which it was hoped this technique might solve, and to cite the complete evidence. These comments would seem to be particularly pertinent to articles appearing in the *Psychological Bulletin*. We have already indicated deficiencies in the citation of evidence. It will, perhaps, clarify the problem if a short history is given here.

It is a truism in rating-literature that halo, leniency, and rater differences in standards are basic problems. The Army Officer Efficiency reporting system, as exemplified in WD AGO Form 67, had become increasingly subject to these influences.[9] Figure 1 illustrates the increasing tendency for officers to be rated higher with the passage of time. Form 67 was not believed by the Army to be serving its purpose, largely because it had lost its discriminating value at the high end of the scale. The Army directed the Personnel Research Section to develop a rating system which would meet its needs to a greater degree. The basis for the research on the problem of efficiency reporting was the study concerned with the development of procedures for the integration of officers into the Regular Army following World War II. This research has been outlined by Richardson (5). The first Personnel Research Section studies of the forced-choice technique were undertaken under this program.[10]

---

[7] The evidence for grade bias, i.e., higher ranking officers being rated higher, in Table II, despite careful instructions to disregard grade, is a case in point.

[8] Before leaving this section, a slight error should be corrected. Travers attributes the quotation beginning "a single over-all rating on a 20-point scale . . . " on page 67, to his reference No. 4. The quotation is actually contained in his reference No. 5.

[9] It is probably more accurate to say that Form 67 had become increasingly subject to halo and leniency. No direct evidence is available concerning variations in rater agreement.

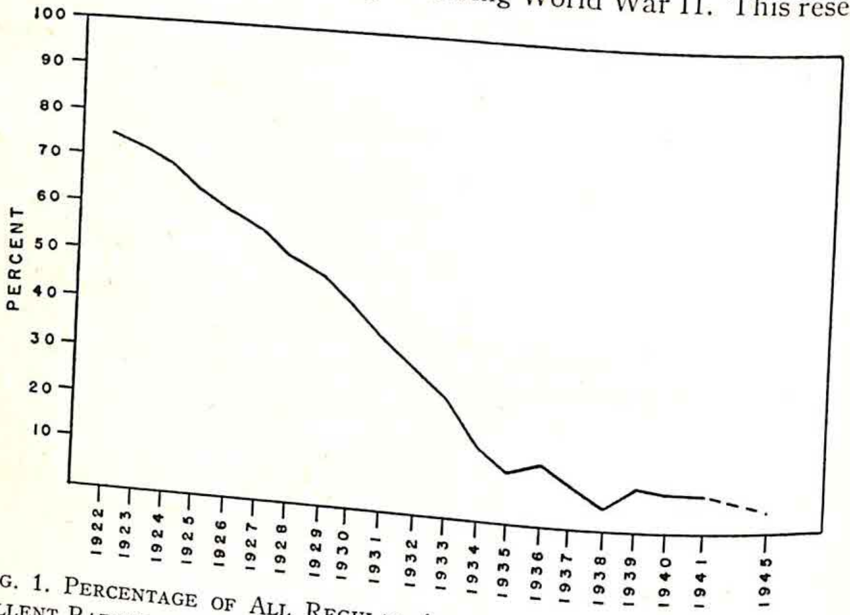

FIG. 1. PERCENTAGE OF ALL REGULAR ARMY CAPTAINS RECEIVING LESS THAN EXCELLENT RATINGS ON WD AGO FORM 67, 1922–1945
No Ratings for War Years 1942–1944.

[10] In the interest of historical accuracy, some elaboration of Travers' statement concerning the origin of the forced-choice technique should be made. The idea of forced-choice was suggested by Dr. Paul Horst in a discussion at an APA meeting. He himself, does not remember the incident. Dr. Wherry was sufficiently interested to develop the scaling methods to achieve the process as Dr. Horst had discussed it; namely, to present, simultaneously, items which looked alike to the individual completing a personality scale and yet had differing significances. Dr. Wherry developed the scaling procedures while working for the Civil Aeronautics Authority and brought them with him when he came to the Personnel Research Section. Jurgensen (3) during this period had been working on a somewhat similar idea, a fact which was not known to the Personnel Research Section until after World War II. It was Wherry's basic scaling technique

Following this integration program, studies were initiated which compared five different rating forms, including rankings, various kinds of traditional-type ratings, and forced-choice. The first series of studies involving some ten thousand officers in the United States and Europe showed a slight superiority for the forced-choice rating form. This superiority, coupled with the hypothesis that there would be less change in the distribution for the form involving the forced-choice items when the form was used on an official basis, led to the decision to use it along with the old Form 67 at the June, 1946, rating period. Details with respect to skew and means are presented elsewhere (22). Results bearing on the validity have already been presented in Table I.

This is not the place to review in detail the further findings of these and subsequent studies. It seems more helpful at this point to summarize the advantages and disadvantages of the forced-choice procedure as applied in the Army efficiency reports.

The principal disadvantage is that the use of the technique had tended to be unacceptable to Army officers (although, apparently, more acceptable in industry). Acceptability is an especially important problem in rating procedures because of the effect on the rater's attitude. Two comments may be made about the unacceptability:

1. The name, "forced-choice," is an unfortunate one. Reasons for its origin are readily understood in the light of the fact that the original presentation asked the individual to pick one of the two items as most descriptive of himself. Actually, as previously suggested, forced-choice might better be considered a scaled multiple-choice item.
2. The second point that should be made concerning the acceptability of the items arises out of the conversion of the raw scores on efficiency reports to standard scores. Converting to a relative scale caused raters to feel that their ratings were not properly represented by a particular standard score, especially those below average. In objecting to Form 67-1, there was much confusion between the effects of the forced-choice technique and the effects of the use of a relative standard score scale.

The advantages of the technique may be summarized as follows:

1. It reduces halo. Raters completing graphic ratings within the same form tend to mark them all pretty much the same way; i.e., the correlation between graphic rating scales is high. In completing two sections of forced-choice items, raters likewise tend to mark them pretty much the same; i.e., forced-choice sections also correlate high, but not as high as rating scales. Forced-choice ratings and rating scales correlate less than do rating scales with rating

---

which served as a point of departure for the work of the Personnel Research Section, first in the application to the development of personality inventories and later in the application to efficiency reporting. As noted in this article, the technique has been most successful in application to self-rating, i.e., personality inventories.

scales, or forced-choice with forced-choice ratings (**17**). On the simple basis of lower intercorrelation, a combination of the two techniques would have greater possibility of increasing validity. In comparison with Form 67, Form 67-1 has persistently yielded slightly greater validity, perhaps for this reason.

2. It reduces bias; for example, it is less influenced by rank of the rater-officer than was Form 67 (**10**). On the whole, the total score on Form 67-1 agrees better with an average criterion rating than do any of its sections (**9**, **17**). The use of the average in itself is a conventional means of reducing bias.

3. Forced-choice item validities tend to be stable over a period of time. From December, 1946, to January, 1949, item validities correlated from .50 to .60 (**8**). This stability is especially noteworthy in view of the narrow range of these item validity coefficients. Although this evidence needs confirmation, it is the sort of evidence which encourages experimentation with the technique.

4. Raters agree better on a report composed of both types of technique than they do on either type alone (**17**). This finding is of first importance. Ratings on Form 67-1 for two successive rating periods for a group of 914 raters showed the best agreement for a combination of both techniques (i.e., total score) than for either technique (Table III).

TABLE III

AGREEMENT OF RATERS ON SUCCESSIVE REPORTS, FORM 67-1

Traditional rating sections:	*r*	*Forced-choice sections:*	*r*		*r*
Section V	.47	Section IV	.42	*Total Score*	.56
Section VII	.39	Section VI	.45		

In the absence of other than a rating criterion, the problem of validity of efficiency reports may reduce to one of this kind of reliability. Thus, the greater rater agreement on a combination of the two techniques is of special significance.

It should be noted that the criterion used for the validity coefficients presented above was the average of a series of rating scales. Rating scales would, therefore, be favored, a point which Travers neglects to mention in his discussion. A further point which is illustrated in the above tabulation is that rating scales and forced-choice ratings may differ among themselves. It is, therefore, difficult to make any hard and fast generalizations concerning either type of rating.

The research of the Personnel Research Section, plus certain theoretical considerations, has persistently affirmed the value of the forced-choice technique. Since in the Army, at least, efficiency reports are usually considered for the entire career and since no technique or combination of techniques has brought rater agreement up to a satisfactory value, our attention has been directed to developing a system whereby fluctuations in rating owing to leniency or other purely biasing factors might be reduced. To put it another way, so much more is gained by combining ratings made by different raters than by improving the rating of a single rater through the use of a special technique that our emphasis is on averaging the reports prior to making use of them. Obviously, for such a system to work, an adequate distribution of single ratings must be maintained. From Figure 1 it is clear that over a period of time, ratings on the traditional type of rating scale in the Army tend to become restricted to the upper portion of the scale. If this kind of trend can be established as characteristic, it would appear to be necessary to develop techniques such as forced-choice which have the promise of maintaining a spread in the ratings.

It should be further noted that in the industrial situation, where people may be rated consistently by the same rater, this averaging system probably will not work. In such situations, therefore, an effort to develop techniques such as forced-choice is particularly needed.

In conclusion, three points should be emphasized: Travers discusses forced-choice technique as applied by itself. It should be observed that both in this reply and in the research reports of the Section, the value of forced-choice in combination with the traditional type of rating scales has been stressed. Until an experiment is set up which gives each technique an equal chance to prove its worth—that is, the correlation not being subject to technique contamination—conclusions as to the value of techniques used will not be definitive.

Secondly, the forced-choice technique has been discussed in terms of the way it has been applied. There are many ways in which it might be improved, e.g., in method of calculating preference and discriminative values, in method of pairing alternatives, perhaps by further application of suppressor theory or better grouping in terms of item content. And, finally, the well-demonstrated value of the technique in the construction of self-rating scales of the personality inventory or self-descriptive type is again stressed.

BIBLIOGRAPHY

1. Brogden, H. E., Newkirk, G. F., & Loeffler, J. C. The prediction of officer potential of ROTC cadets. *Amer. Psychologist*, 1950, 5, 360. (Abstract.)
2. Campbell, J. T., & Rundquist, E. A. Scale items for inclusion in forced-choice rating form. *Amer. Psychologist*, 1950, 5, 280. (Abstract.)
3. Jurgensen, C. E. Report on the "Classification Inventory," a personality test for industrial use. *J. appl. Psychol.*, 1944, 28, 445–460.
4. Karcher, E. K., & King, S. H. Effect of number and order of ratings on reliability and validity. *Amer. Psychologist*, 1950, 5, 333. (Abstract.)
5. Richardson, M. W. Selection of Army officers. In *New methods in applied psychology*, Proceedings, Maryland Conference on Military

Psychology. Univ. of Maryland, 1947, 79–85.

6. RICHARDSON, M. W. An empirical study of the forced-choice performance report. *Amer. Psychologist*, 1949, 4, 278–279. (Abstract.)
7. RUNDQUIST, E. A. Personality tests and prediction. In D. H. Fryer and E. R. Henry, *Handbook of applied psychology*. New York: Rinehart, 1950. Vol. 1, 182–191.
8. RUNDQUIST, E. A., WINER, B. J., & FALK, G. H. Follow-up validation of forced-choice items of the Army Officer Efficiency Report. *Amer. Psychologist*, 1950, 5, 359. (Abstract.)
9. SCHNEIDER, DOROTHY E., & BLACKBURN, J. R. Validity of a graphic scale of officer efficiency. *Amer. Psychologist*, 1950, 5, 359. (Abstract.)
10. SISSON, E. D. Forced-choice—the new Army rating. *Personnel Psychol.*, 1948, 1, 365–381.
11. STAFF, PERSONNEL RESEARCH SECTION, AGO. The forced-choice technique and rating scales. *Amer. Psychologist*, 1946, 1, 267. (Abstract.)
12. STAFF, PERSONNEL RESEARCH SECTION, AGO. Construction and selection of items for the Biographical Information Blank (BIB). PRS Report No. 703, 7 July 1945.
13. STAFF, PERSONNEL RESEARCH SECTION, AGO. Validation of Form E of the Biographical Information Blank. PRS Report No. 716, 8 July 1946.
14. STAFF, PERSONNEL RESEARCH SECTION, AGO. Comparison of the rating check list (RCL) and forced-choice list (FCL) methods of obtaining ratings. PRS Report No. 717, 9 July 1946.
15. STAFF, PERSONNEL RESEARCH SECTION, AGO. A comparison of officer efficiency ratings obtained with the WD AGO Form 67 and the WD AGO Form 67-1 at the Regular Reporting Period of 30 June 1946. PRS Report No. 725, 17 Jan. 1947.
16. STAFF, PERSONNEL RESEARCH SECTION, AGO. Item analysis and development of scoring keys for the Leaders' Course BIB. PRS Report No. 764, 1948.
17. STAFF, PERSONNEL RESEARCH SECTION, AGO. 1949 Follow-up validation of WD AGO Form 67-1. PRS Report No. 791. (In preparation.)
18. STAFF, PERSONNEL RESEARCH SECTION, AGO. Determination of preference and discrimination values of phrases for preference check list sections of alternate forms of Officer Efficiency Report, WD AGO Form 67-1. PRS Report No. 846. (In preparation.)
19. STAFF, PERSONNEL RESEARCH SECTION, AGO. Development and validation of the Self-Description Blank for predicting leadership qualities of ROTC cadets. PRS Report No. 859. (In preparation.)
20. STAFF, PERSONNEL RESEARCH SECTION, AGO. A rationale for minimizing distortion in personality questionnaire keys. PRS Report No. 868. (In preparation.)
21. TAYLOR, E. K., CARROLL, J. B., & WINER, B. J. Validity of the Army's officer efficiency report. *Amer. Psychologist*, 1949, 4, 284. (Abstract.)
22. TAYLOR, E. K., & WHERRY, R. W. A study of leniency in two rating systems. *Personnel Psychol.*, 1951, 4, 39–47.
23. TRAVERS, R. M. W. Validity and rationale of forced-choice technique. *Psychol. Bull.*, 1951, **48**, 62–70.

Received April 19, 1951.

NOTE ON TRAVERS' CRITICAL REVIEW OF THE FORCED-CHOICE TECHNIQUE[1]

MARION W. RICHARDSON
Richardson, Bellows, Henry and Co., Inc., New York

With reference to the hesitation of Travers "to accept the evidence provided by that author [Richardson] and his associates since the procedure involved seemed to raise spuriously correlations between assessments of job performance based on a forced-choice scale and an independent criterion of job proficiency," Richardson pleads *mea culpa* in only one respect. The reviewer did not have access to enough data, because of the slowness of publication. That is our fault.

Once the question came up as to the evaluation of psychological research performed for industrial clients by agencies such as Richardson, Bellows, Henry and Co., and my suggestion was that any research results reported by us would be referred willingly by us to a competent competitor such as the Psychological Corporation. I am confident that the Psychological Corporation would have similar reactions. In this situation, we should give Dr. Travers the opportunity to examine the basic evidence both in the raw and in the analyzed form. It does not necessarily follow that because new techniques are being tried out in a nonacademic environment, the procedures are unsound.

I will try to suggest some of the things the reviewer will discover if he accepts the invitation or when RBH properly gets around to publication.

First, let us dispose of the purified criterion issue by some generalizations no more broad than the reviewer's.

1. *The purified criterion technique is not a necessary feature of the forced-choice technique.* It has not always been used in the thirty odd performance reports developed by our firm.

2. *The technique probably has the effect of reducing the dimensionality of the behaviorial complex we call "effectiveness on the job."* If true, the effect is to simplify the measurement problem by eliminating some of the variance due to the more elusive factors. Surely it is legitimate to decide whether one does the simpler job reliably or the more complicated one unreliably.

3. *If one applies the scale to the entire sample population (thus using the unpurified criterion), the shrinkage in the correlation coefficient is not so large as Travers imagines.*

In some of our studies the purified criterion is used to select the original items for inclusion in the groups (blocks of three, four, or five), and, again, to select blocks from the total number tried out. (The latter procedure is not mentioned by the reviewer, since Sisson did not list it.) Although the validity coefficient may be computed on the purified criterion, it has always, or nearly always, been recomputed on the entire sample in a manner that ought to satisfy the statistical purist's soul. In our studies the shrinkage has not been alarming, but Travers has a right to see the evidence in forthcoming publications, or otherwise.

---

[1] TRAVERS, R. M. W. A critical review of the validity and rationale of the forced-choice technique. *Psychol. Bull.*, 1951, **48**, 62–70.

The real statistical issue is the shrinkage of validity coefficients on successive samples drawn from the same population. This is the *bête noir* of most validity studies. The one scale criticized by the reviewer was applied by a competent independent investigator to a new group not used for standardization of the scale.

M. Trawick of the Esso Standard Oil Company reports on the same scale, a validity coefficient of $r=0.76$. A key official of the Standard Oil (N.J.) Company has (as yet unpublished) résumés of several repeat studies on the scale criticized by Travers. With the purified criterion issue completely absent, and on a background of two years' experience with the scales under rigorous tryouts in new situations, the validity coefficients ranged from $r=0.60$ upward, and several exceeded (within limits of standard error) the original coefficient reported. Under the conditions set, there was no possibility of criterion contamination.

We are much more interested in such tests, some of them on new samples with mixed cultural and linguistic groups, than we are on a minor statistical point such as the reviewer brought up. We can at least retort "guilty in part, but what of it?"

This amount of correlation shrinkage has not, to date, worried our investigators. We do not know whether the use of the purified criterion technique increases, leaves unchanged, or decreases the all-important shrinkage on new samples. It may be argued on an arm-chair, but inconclusive, basis that the *important* shrinkage is actually decreased, but more evidence is needed.

The present writer is inclined to agree that further evidence on negative skew and elimination of control of final rating by the rater is needed.

If the criticized may turn on the critic, I would express the opinion that the "rationale" portions of the article are too fragmentary and incomplete. A proper rationale would be mathematical, involving a "difference" function the present writer has not yet been able to handle.

The reviewer's one constructive suggestion seems to involve directions too complicated to be carried out on ordinary people, and in its feasible special case, to lead us right back to the technique he doesn't like. For example, how much difference is there between *forcing* raters to give a different rating to every trait, and *forcing* a choice of one of two or more traits as most characteristic?

Again, the writer pleads *mea culpa* only with respect to his failure and that of his associates to furnish data on the now-numerous applications of the so-called forced-choice techniques. (There is no "stock-in-trade forced-choice instrument," as averred by the reviewer.)

Received April 19, 1951.

BOOK REVIEWS

Mowrer, O. Hobart. *Learning theory and personality dynamics.* New York: Ronald, 1950. Pp. xviii+776. $7.50.

Mowrer has taken a leading role over the last few years in providing integration at the level of theory between the two fields suggested by the book's title: the laboratory studies of learning, and the clinical study of the person. The struggle upon which he has been engaged calls for a reinterpretation and reconciliation of the viewpoints stemming from Freud and Pavlov.

The book represents selections from the author's earlier papers, supplemented by a good deal of previously unpublished material. Following an introductory chapter, the book is divided into two parts. The first part, entitled "Learning Theory," consists of twelve chapters, of which three contain previously unpublished material. The second part, "Personality Dynamics," also twelve chapters, includes four new ones. The previously published chapters are supplemented by footnotes to indicate the author's changes in point of view since the papers first appeared.

It will be convenient to permit remarks about the book to correspond to the two parts. First we shall consider the contributions to learning theory, followed by attention to what the author has to say about personality.

Learning theory. The two major contributions to learning theory are, first, the conception of acquired fear as a secondary drive, and second, a dual theory of learning, distinguishing between conditioning (simple associative learning) and problem-solving (learning with reinforcement, under the law of effect).

That fear-reduction may be reinforcing is now quite widely accepted by learning theorists whose points of view are related to those of Hull. Mowrer deserves credit for originally proposing this conception, and for building the apparatus (with Miller) that has served to provide convenient laboratory demonstrations of the possibility of reinforcement through escape from noxious stimuli, and, in later modifications, through escape from fear.

Accepting fear as a learned drive is not the same as accepting fear as the *only* learned drive, although Mowrer at points comes near to doing this. Impressed by the role of fear in learning through punishment, he finds that some experiments on food or water deprivation may also be interpreted through the mediation of fear. Such interpretations are offered of an experiment by O'Kelly and Heyer (p. 347) and one of Estes (p. 349).

This is not the place to review these interpretations in detail, but the reader has to be warned that the treatments are not entirely self-consistent, and such inconsistency as remains is not solely a matter of the range of dates covered by the original papers. Mowrer's greatest difficulty comes in trying to explain secondary reinforcement by a stimulus associated with reward. A familiar example is the reward-value of the click associated with the food-delivering mechanism in a Skinner Box. In a passage inconsistent with his fear-reduction theory, Mowrer has this to say of the secondary reinforcing effect of a signal produced by the organism's response: "However, what is important here is merely that the signal take on a *pleasant connotation*, so that when the subject happens to produce the signal, the response which was effective in producing it will be reinforced" (p. 323).

Psychoanalytic theory gives a central place to anxiety-reducing mechanisms. The familiar defense mechanisms are said to serve this purpose. To show the operation of related principles at the rat-learning level has provided a useful bridge between animal learning and psychoanalytic theory. This much we may accept, without accepting fear as the only learned drive, and recognizing with Mowrer that the rat's fear is not the same as the human being's anxiety. We owe Mowrer a debt for initiating and advancing this line of thinking.

The second feature of learning theory, stressed over and over again in the book, is the need for a dual theory. Mowrer has rather reluctantly joined the ranks of the many earlier writers who made a distinction in kind between simple associative learning and rewarded learning. He attempted earlier, with Hull, to reduce all learning to the paradigm of reinforcement. But now he is convinced that the original effort was faulty. His footnotes, reinterpreting earlier results, and many supplementary discussions, seek to make a case for the modified position.

Conditioning (the first of the two kinds of learning) is "the process whereby emotional learning . . . takes place" (p. 236). "It now seems preferable to apply the term 'conditioning' to that and only that type of learning whereby *emotional* (visceral and vascular) responses are acquired . . . Many responses which have previously been termed 'conditioned responses,' are in the present conceptual scheme, not conditioned responses at all. Only those responses which involve visceral and vascular tissue and are experienced subjectively as emotion are assumed to be conditioned responses" (p. 244).

Problem-solving (the second of the two kinds of learning), is equated with effect learning. "Effect learning has been previously conceived as applying mainly to those situations in which the motive, or 'problem,' is an unlearned biological drive, such as hunger, thirst, pain, etc. It is now clear that effect learning may be expanded to include those situations in which the motive, or 'problem,' is a *learned* drive, i.e., an emotion such as fear or an appetite. . . . If an emotion, or secondary drive causes the skeletal musculature to be activated, and if such activity results in secondary drive reduction, then the overt response thus acquired is here conceived as an instance of effect learning, not conditioning" (p. 244).

The problem that gave rise to this reinterpretation was that of how the acquired fear came about that was to serve as the secondary drive, and hence provide reinforcement when reduced. Mowrer has decided that the solution is to accept a simple association between the fear responses and the incidental stimuli present at the time that fear responses occur. Hence a box in which an animal is shocked (and frightened) will later produce fear even in the absence of shock. This fear serves as a drive to motivate learning that will reduce it.

The choice that Mowrer felt called upon to make (and that he invites his reader to make) is between a monistic reinforcement theory and his dual theory. There are many more alternatives. Mowrer makes a pretty strong case against a monistic reinforcement theory, but this by no means leaves his dual theory as the only (or even the preferred) alternative.

Others who accept simple contiguity theories (e.g., Thorndike's associative shifting, and Guthrie's conditioning) do not find it necessary to limit such associative learning to emotional responses. The evidence for so limiting associative learning is no better than for extending it, say, to word-association learning.

The criticisms of the monistic reinforcement theories are to be taken seriously. The dual theory that Mowrer proposes has to be examined as one of the alternatives, but it is not the only one, and needs much tightening to become a strong candidate for wide acceptance.

Personality dynamics. One is led to expect a much clearer integration between learning theory and personality dynamics than he actually finds.

The duality theory of learning is said to have its counterpart in Freud's distinction between the reality principle and the pleasure principle. To be sure, the pleasure-principle, concerned as it is with immediate gratification and tension-reduction, sounds like the law of effect. But does this, then, make the reality principle correspond to primitive emotional conditioning? In Freudian theory, the reality principle comes in later on, as the organism learns to postpone gratification in order to deny the painful consequences of seeking immediate pleasure. This is problem-solving; it is not simple conditioning. Mowrer might have toyed with Freud's repetition principle as coming nearer to simple associative habit. But this possible analogy he did not follow.

The treatment of personality dynamics arises much more out of reflection upon Freud's theory as it applies to clinical data than out of inferences based on learning theory. The main emphasis is upon disagreements with Freud. This is interesting because Mowrer was one of those who, in earlier writings, helped make Freudian theory plausible to experimentally oriented psychologists. The somewhat reluctant disagreement with Freud parallels in a way the disagreement with reinforcement theory.

As specimens of the disagreement we may consider the theory of identification, the theory of neurosis, and the theory of anxiety.

Mowrer believes that both boy and girl achieve an identification with the mother, before sex-roles are differentiated. The sex-role identification is then with the like-sexed parent, coming later. This is contrasted with the Freudian theory of a first choice on a sexual basis ("object-choice") of the opposite-sexed parent, followed later by identification.

Mowrer accepts an immaturity theory of neurosis, as contrasted with an overlearning theory that he attributes to Freud. That is, neurosis for Mowrer signifies an underlearning of what is needed to get along, while neurosis to Freud often means the need to unlearn the prohibitions that were too well learned in childhood. It seems somewhat doubtful that this is fair to Freud, and the strong possibility exists that both kinds of neuroses may be found. In Freudian terminology, some cases have an underdeveloped superego ("holes in the superego"), while others have an overwhelming superego.

The third difference with Freud lies in Mowrer's "guilt" theory of anxiety as contrasted with Freud's "impulse" theory. In Mowrer's words:

"In essence, Freud's theory holds that anxiety comes from evil wishes, from acts which the individual would commit if he dared. *The alternative view here proposed is that anxiety comes, not from acts which the individual would commit but dares not, but from acts which he has committed but wishes that he had not. It is, in other words, a 'guilt theory' of anxiety rather than an 'impulse theory'* " (p. 537).

There are consequences for therapy arising out of these differences with Freud. In general, the relief of the pressure of guilt through confession leads to a kind of conversion experience. Neurotic anxiety becomes normal anxiety, and new learning can occur. Therapy does not consist in watering down the superego, but in growing up to the assumption of responsibility. "Valid treatment should . . . lie in the direction of helping the individual to grow up, emotionally and socially, to the point where the demands of conscience and community are understandable and acceptable" (p. 572).

We have here a very provocative book, certainly alive to the issues that are being debated in contemporary psychology. The book is difficult, but it is thoughtful, and the careful reader can learn a great deal from it.

It is to be regretted that the author did not make the effort to rewrite his papers instead of merely annotating them. The volume as it is offered the reader is more difficult than it needed to be, for it carries along all the asides and preoccupations of the author as he was feeling his way toward his final theory. A much smaller book might have carried his message more clearly to a wider audience.

Stanford University. ERNEST R. HILGARD.
University of Michigan. EDWARD L. WALKER.

VÖLGYESI, FRANZ ANDREAS. *Hypnosetherapie und psychosomatische Probleme.* (Hypnotherapy and psychosomatic problems.) Stuttgart: Hippokrates-Verlag Marquardt & Cie., 1950. Pp. 203. DM 8.25.

There would be little purpose in reviewing a book that is very unlikely to come to the attention of most American psychologists and psychiatrists, were it not for the fact that, at least according to the publishers (Germany, U.S. Zone), the author's pronouncements aroused great interest among the readers of a German medical journal (*Hippokrates*), and, far more important, that the book gives a very significant glimpse of the "progress" of psychological science behind the Iron Curtain. The ominous suffusion of scientific work with political ideology of which this book—probably against the author's intentions—gives eloquent evidence, is sufficient reason for Western readers to take a good look at the product and shudder at the perversions of science under a dictatorship: from there it is but a short step to the abominations perpetrated by the Nazis, when calculated sadism and utter disregard for human dignity paraded under the guise of "research."

Völgyesi is a Hungarian medical researcher who has written extensively on the subject of hypnotherapy and reflexology, deriving his theoretical inspiration from Pavlov, Bekterev, and their students, whose work is cited throughout the book. Reflexology, then, is the conceptual framework for the author's hypnotherapy, the aim of which is "a general, central reorientation of the patient by means of psychocortical, cortico-visceral, psychosomatic, suggestive effects, which are determined by conditioned reflexes" (Foreword, p. 7). This becomes clearer in chapter II, which is entitled "So-Called 'Psychosomatic Medicine.'" It is admitted there that "Anglo-Saxon" psychosomatic medicine has had a fruitful influence, but certain "essential corrections" have to be made. These encompass (1) the historical side of the psychosomatic viewpoint, (2) the psychoanalytic overemphasis, and (3) the "so-called 'independent psychosomatic diseases.'" The main line of attack is directed against the psychoanalytic orientation of psychosomatic medicine in this country, and the "favorite attempt" to postulate "independent psychosomatic syndromes." We learn that the latter do not exist, just as there is no "purely somatic" or "purely mental" illness. In every disease, both aspects are said to be inextricably interwoven. Quite apart from the fact that the foregoing is a gross misrepresentation of present-day thinking in psychosomatic medicine, it is more instructive to be told that since the time of Aristotle "*body and soul have formed a social-dialectic*, that is, *antinomical entity*." (All italics are the author's.)

A great many examples are cited to demonstrate that "*active hypno-suggestive psychotherapy achieves its results despite somatic, toxic, etc. disturbances*" (p. 147). And, quoting from Grastchenkov, "*the relationships between etiology and pathogenesis cannot possibly be treated scientifically without the principles of reflexology*" (p. 102). Asserting that all neuroses and all types of neurotic manifestations always have organic bases, Völgyesi attempts to establish "dialectically interpreted" hypnotherapy as a technique which supposedly produces astounding cures for a variety of medical and psychiatric problems. But this is not all; not only is hypnotherapy *the* psychotherapy par excellence (psychoanalysis is alternately termed "reactionary" and an "aberration"), but it also produces structural changes in the organism's repertoire of unconditioned reflexes. Thus, it somehow affects the hereditary characteristics of the organism. On page 91 we read: "Suggestion and conditioned reflexes are capable of altering, with experimental exactness, the hereditary unconditioned reflexes."

If these were one man's views, they could easily be disregarded. But the multitude of experimental investigations cited show with incontrovertible clarity that this represents "official" Soviet thinking:

> The final summary of the joint serial sessions of the Scientific Academy and the Medical Scientific Academy of the Soviet Union, in June 1950, made serious reproaches (*erhob schwere Vorwürfe*) against Orbeli, Beritasvili, Anochin, Kupalov, even against Speransky, and their schools: despite all of their valuable achievements relative to conditioned reflexology, they did not follow sufficiently *the original concepts and the methodology of Pavlov's doctrines* regarding nervism and materialistic neuro-psychiatry, and the concluding note of the final summary "*calls upon every worker in physiology and medical science, to promote* in constructive fashion, with free criticism and self-criticism, *the great doctrine of Pavlov, for the benefit of the people*" (footnote 5, p. 104).

Implicit in these formulations are (1) an assumption concerning an almost limitless modifiability of the human organism, which makes it capable of being directly influenced by suggestion, (2) a paradoxical rigidity once modifications have been made through suggestion, such that the hereditary structure is permanently altered, and (3) a fanatical belief in the all-inclusiveness of conditioned reflex principles as an explanatory construct. Irrespective of the scientific truth of the theory, which is at best partial and certainly contravenes much evidence available today, this book is a glaring example of what happens when research is built around political ideology and when experimental findings are interpreted after the scientist has made certain that they fit in with "official" thinking. May this never come to pass in American psychology!

HANS H. STRUPP.

Human Resources Research Laboratories,
Bolling Air Force Base.

ROSENBLUETH, ARTURO. *The transmission of nerve impulses at neuro-effector junctions and peripheral synapses.* New York: Wiley, 1950. Pp. xiv+325. $6.00.

Part I of this monograph stems from the 1937 work by Cannon and Rosenblueth, *Autonomic Neuro-Effector Systems.* It is concerned with the liberation of chemical mediators by nerve endings at their junctions with smooth muscle, cardiac muscle, and glands. It reviews the evidence that acetylcholine is the substance liberated by the cholinergic axons located chiefly in the parasympathetic system. The controversial problem of the adrenergic axons is dealt with by supposing that either adrenaline or an "adrenaline-like substance" is released at their endings, which occur chiefly in the sympathetic system. The action of the adrenaline-like substance is further supposed to be conditional upon the presence of a receptive substance at the junction with the effector; it is the excitatory or inhibitory nature of this substance which results in the formation by adrenaline of one of two hypothetical substances, sympathin E and sympathin I. Rosenblueth thus maintains in this work the principal contention of Cannon and Rosenblueth in their 1937 monograph: that the adrenergic effects are those of a mediator combined with a local agent whose function is to regulate the excitatory or inhibitory response of any particular effector at any given time. While there is some discussion of recent experiments in that area, the present treatment contains no change in basic point of view.

Part II has to do with peripheral synapses, broadly conceived to include not only the neuro-neuronal junctions in autonomic ganglia but also the junctions of somatic motor nerve fibers with striated muscle. The author states, "The argument for chemical transmission at peripheral synapses appears stronger than that which can be made for chemical transmission at autonomic neuroeffector junctions, yet the latter transmission is generally accepted as chemical whereas the former is still considered electrical by many of the experts in the field." Rosenblueth cites as evidence in support of chemical transmission experiments on Wedensky inhibition, Wallerian degeneration, and the Philipeaux-Vulpian phenomenon. He considers the effects of curare, veratrine, and potassium ions, and assigns to the latter the function of adjuvant to acetylcholine, the principal mediator. The spike potential is regarded as a mere manifestation of a depolarization of nerve membrane which is assumed to be the condition essential for the release of acetylcholine.

Incidentally, there is no mention in the book of the ideas of Dr. Norbert Wiener, whose book, *Cybernetics*, contains a tribute to Rosenblueth for his understanding of scientific methodology.

However well presented, the thesis of this book will no doubt be widely challenged, and Rosenblueth will be criticized for clinging too tenaciously to theories advanced more than a dozen years ago by Cannon and himself. This he does in spite of recent significant work of Bacq, Bozler, Gaddum, Feldberg and others. Rosenblueth does take cognizance of this newer work, notably in chapter IV, but it is apparent that it has had little effect in modifying his interpretation of the role of electrical events, to cite a specific example, in smooth muscle contraction. J. H. Gaddum some time ago remarked drily that there could be no doubt of the importance of the experiments by Cannon, Rosenblueth, and others which led to the theory of sympathins E and I, but that the theory itself should be forgotten. It has been Gaddum's contention that no known substance has the properties ascribed to sympathins E and I, whereas all the experimental observations can be accounted for on the basis of known substances, notably adrenaline and noradrenaline.

Another authority in this field, Z. M. Bacq, is of the opinion that acetylcholine, adrenaline, and noradrenaline are the "local hormones" or chemical mediators which are clearly established at the present time. He states that there is universal agreement among English, Swedish, and Belgian pharmacologists on the point that sympathin is a mixture, in varible proportions, of adrenaline and noradrenaline. The confident enthusiasm of Rosenblueth may be contrasted with some recent observations of G. L. Brown to the effect that we are perhaps within sight of an explanation of the mode of action of acetylcholine, and we are in a similar state of obscurity with respect to the action of adrenaline as a chemical mediator.

The reviewer, no expert in these matters, has talked with Dr. F. G. Sherman of Brown University and other colleagues among the physiologists. He has received the impression that Rosenblueth's monograph will be received by serious students as an excellent compilation and introduction to the difficult literature on chemical mediators. Yet there is a distinct tendency to go along with the authorities enumerated above in rejecting hypothetical mediators in favor of those whose properties and chemical structure, at least *in vitro*, are well defined.

LORRIN A. RIGGS.

Brown University.

GOULDNER, ALVIN W. (Ed.) *Studies in leadership*. New York: Harper, 1950. Pp. xvi+736. $5.00.

Studies in Leadership is a collection of thirty-two papers representing the work of thirty-five different authors. Twelve of the selections have been published previously, while twenty are new contributions. After an introduction by Gouldner the papers are organized into five groups. Part one, "Types of Leaders," deals with informal leaders, bureaucrats, and agitators. "Leadership and Its Group Setting" is the title of the second part, which is concerned with leadership in various social classes, in several minority groups, and in diversified political atmospheres. Part three, bearing the provocative title "Authoritarian and Democratic Leaders," deals with problems of democratic, manipulatory, and authoritarian leadership. Part four is called "The Ethics and Techniques of Leadership," and the descriptive title of the final group is "Affirmations and Resolutions."

Something of the tone of the volume can be gathered from an examination of the headings given above and even more from a consideration of the method and scientific orientation of the articles. According to the reviewer's classification, four of the papers are largely historical. Typical of such articles is Cox's "Leadership among Negroes in the United States" or Green and Melnick's article on the feminist movement. Seven contributions are oriented around material gathered through the field observation of some leader or social group. Some of Leighton's important conclusions, based on material gathered in the relocation camp for Japanese at Poston, are presented. There is an interesting case study of a local union leader by Alexander and Berger. Also included are three short selections from Whyte's *Street Corner Society*. Feuer presents some stimulating observations on the leadership and organization of collective settlements in Israel. Another seven studies are more "data oriented" in the sense that considerable statistical material is given as a basis for the presentation. Typical here is material from "The People's Choice." Particularly interesting is a study by Lipset on the leadership of the Cooperative Commonwealth Federation (CCF) of Saskatchewan. Finally, there is a group of fourteen contributions, or almost half the articles, which are predominantly speculative, philosophical, or theoretical. While a number of the essays in this group are stimulating, provocative, and insightful, many of them are not scientific in the sense that the authors do not make use of any reasonably explicit body of factual material to serve as a basis for their theorizing.

If Gouldner's value judgments of the good in leadership work are typical of sociologists, and if these papers fairly represent such values, then a considerable gap in opinion regarding desirable methodology exists between those working in sociology and those in psychology. The reviewer feels that most psychologists are wedded to an approach to leadership problems (and social psychological problems generally) which is "fact oriented." While theory is needed, it is not allowed to run into the broadest generalizations unchecked by data. One of the authors in this collection states, "Almost the entire literature on leadership stems in large measure from the writings of Aristotle and Machiavelli" (p. 396). Indeed, here would seem to be the major criticism of many of the selections: while interesting, they are essentially social philosophy, which, if this book is evidence, is not based on any comprehensive set of data.

Such considerations raise the question as to the selection of material. In such a fast-growing field as leadership it is difficult to keep abreast of all the newest developments, but it is disappointing not to find more contributions from recent psychologically-oriented leadership research. There is nothing from the Ohio State group (Shartle, Stogdill, Hemphill). Cattell's work on syntality is not mentioned. The new developments in leadership and group behavior associated with the Research Center for Group Dynamics is inadequately portrayed by one of Lewin's early (1939) papers. (Incidentally, two of Lewin's papers are reprinted only in part and are given new titles, but neither change is indicated by the editor.) The wartime developments in leadership and group behavior of the Office of Strategic Services and of the British are represented in an article by Eaton, which has been outdated by *The Assessment of Men* and by Harris' *The Group Approach to Leadership-Testing.* Because of such omissions or inadequate representation, many psychologists will feel that *Studies in Leadership* does not give an adequate picture of current work in this field.

LAUNOR F. CARTER.

University of Rochester.

CANTRIL, HADLEY. (Ed.) *Public opinion 1935–1946.* (Prepared by Mildred Strunk.) Princeton, N. J.: Princeton Univ. Press, 1951. Pp. lix +1191. $25.00.

Not only social scientists, but also government officials, business executives, and labor leaders have many occasions for inquiring as to the state of public opinion on a given topic. While this need is likely to be most acute as regards contemporary issues, the significance of today's poll can best be evaluated if it is placed in the context of past observations. The practical man is likely to be particularly concerned with trends: Are Americans becoming more isolationist? Is the public more hostile to labor unions?

All of us who are working with such questions will owe a real debt to Mrs. Mildred Strunk, who did most of the work in compiling this impressive reference volume. She has gathered results from 23 polling organizations in 16 countries for the 11 years from 1935 to 1946, and organized them in an easily consulted format. Evaluation is facilitated by indicating type of sample, percentage of "don't know" responses, and other significant information.

It is to be hoped that Dr. Cantril's organization will carry out its plan for issuing such volumes at five-year intervals. Such a series would have great value for social psychology all over the world.

University of Illinois. ROSS STAGNER.

BRITT, STEUART HENDERSON. *Selected readings in social psychology.* New York: Rinehart, 1950. Pp. xvi+507. $2.00.

The maturity of social psychology—like that of any science—will be marked by more systematic treatment of the kind which provides empirically testable hypotheses. While the beginnings of such maturity are dimly seen in some recent books, we are far from any really satisfactory theory-building and probably will be for some time to come. In the meantime social psychology seems to be in for another spate of symposia and books of readings. In fact, there appear to be certain recurrent fashions in these matters. Perhaps in a rapidly growing and still vaguely bounded field we may expect this cyclic phenomenon.

The vague and still unstructured nature of systematic social psychology is evident in Britt's book—as it is in contemporary textbooks generally—in the variation in major topics presented as well as in the content of the individual papers. Britt has organized his selections into five major divisions: (I) "Social Psychology and Its Methods"; (II) "Biological and Social Foundations of Behavior"; (III) "Some Individual Factors of Social Adjustment"; (IV) "Behavior in the Presence of Others"; (V) "The Social Psychology of Institutions"; and (VI) "Social Conflicts."

As is bound to be true of all such books of readings, the decisions of a particular compiler as to what to include or exclude seldom completely agree with the judgments of others. On the whole this reviewer feels that Britt has done a reasonably good job, though his materials are chiefly of contemporary character. Certainly some of the papers are of such ephemeral value that they will quickly go out of date. Personally I should like to have seen some of the classical, though admittedly theoretical, papers included. For example, pertinent selections from James, Cooley, Dewey, George H. Mead, and Lippmann would provide the student-reader with an historical and somewhat systematic background.

In a review of this character one cannot summarize or criticize the whole range of the fifty selections which make up this book. In keeping with present trends Britt recognizes the importance of social-cultural materials for social psychology. Papers by Linton, Warner, La Barre, and Withers testify to this fact. The attention given to the topics of social class and social conflict, as is true in the widely used *Readings* by Newcomb and Hartley, is evidence of a growing awareness of the importance of these matters in our everyday life. So far, however, no one has come up with any very testable theory about the function of these factors in relation either to personality or group structure. Since the editor of this volume is professionally much concerned with problems of mass communication, I should have expected more extensive and somewhat different selections on this topic. In this connection a most serious and obvious gap is the failure to cover the field of language adequately. The short selection from Albig is good as far as it goes, but the inclusion of papers on the symbolic process as related to communication, cross-cultural materials (such as from Lee and Whorf), and some of the recent publications of Merton, Lazarsfeld, and their co-workers would have strengthened the volume.

But to this reviewer there is a more serious omission. The usefulness of books of this kind, usually designed for orientation and information to the beginning student, is greatly enhanced when the editor uses interpretative summaries (or introductions) as a method of tying the materials together. Moreover, when such a job is well done, it may also serve to contribute to the growth of systematic theory and thus advance our field at least a little bit. Aside from Britt's own opening paper on methods, originally published in 1937, no effort is made to give any systematic interpretation to the field.

KIMBALL YOUNG.

Northwestern University.

HOMANS, GEORGE C. *The human group.* New York: Harcourt, Brace, 1950. Pp. xxviii+484. $6.00.

Students of social behavior, whatever the label on their academic credentials, should be interested in this provocative book on the face-to-face or primary group. Vigorously positivistic and empirical though neither experimental nor quantitative in his approach, Homans develops in connection with a series of case studies a minimum array of concepts and hypotheses for the analysis of groups as emergent and self-regulatory systems. Indeed, his major contribution seems to me to lie in spelling out what is implicit in this conception of the group as a system, in terms of explicit relationships among analytically distinguished variables. His approach is rooted particularly in the work of L. J. Henderson, Elton Mayo, Chester Barnard, and Conrad Arensberg.

Activity, *interaction*, and *sentiment* are the three primary elements of social behavior with which Homans starts his analysis. While he adds a fourth basic element, *norms*, the mutually interdependent relations among the first three provide the organizing framework for his analysis of group phenomena. In what appears to me as a useful convention, he distinguishes between the *external system* of the group—"the state of these elements and of their interrelations, so far as it constitutes a solution of the problem: How shall the group survive in its environment?"—and the *internal system*—"the elaboration of group behavior that simultaneously arises out of the external system and reacts upon it." In the external system, the problem of motivation involves the relations between sentiment and activity, while functional specialization and the division of labor involve the mutual dependence of activity and interaction. Each of the possible pairings of variables is considered in the analysis of the internal system. Here are examples of the kinds of hypotheses he proposes under each heading:

> Other things equal, contact breeds liking and vice versa. (Interaction and sentiment.)
>
> Persons who feel sentiments of liking for one another will express those sentiments in activities over and above the activities of the external system, and these activities may further strengthen the sentiments of liking. (Activity and sentiment.)
>
> Activities in addition to those of the external system tend to arise among persons who interact regularly, and in turn tend to strengthen their tendency to interact with one another. (Interaction and activity.)

With this conceptual equipment Homans treats in turn clique differentiation, social ranking, leadership, interpersonal relations, social control, and social change.

Five major case studies, each first presented with a minimum of conceptualization but selected to bring out some major aspect of his system, are a notable feature of the book. Faithful to the primary sources, the presentation is a far cry from the perfunctory illustration more frequent in textbooks. The Bank Wiring Room (from Roethlesberger and Dickson's *Management and the Worker*) is used to develop the broad outlines of the internal and external systems. Leadership is considered in relation to the Norton Street Gang (from Whyte's *Street Corner Society*), while the network of interpersonal relations in a family and kinship system is studied with material from Firth's *We the Tikopia*. "Hilltown," a New England community, provides a case of social disintegration (drawn from an unpublished thesis by D. L. Hatch), while a study of industrial morale in the "Electrical Equipment Company" (by C. M. Arensberg and D. MacGregor) serves as a focus for the consideration of social conflict.

I was much struck by the agility with which Homans manages to analyze the relations between particular aspects of behavior in the group without falsifying their interdependence with the entire context of the group as a system. I think social psychologists can study with profit his approach to system, "function," and equilibrium in groups. But his analytical variables seem to me too crude to comprise many of the sensible things he has to say. For example, in discussing leadership he hypothesizes that "*when two persons interact with one another, the more frequently one of the two originates interaction for the other, the stronger will be the latter's sentiment of respect (or hostility) toward him, and the more nearly will the frequency of interaction be kept to the amount characteristic of the external system*" (p. 247, his italics). Yet it is quite clear in his discussion here and elsewhere (cf. esp. pp. 182, 418) that what counts is not *who* originates the interaction, but the *nature* of the interaction originated. A leader or person in authority is identified by the *kind* of interaction he receives *and* originates, not by the gross amount originated and received. In other words, interaction is too gross a concept for the demands Homans places on it. Similar strictures will occur to most psychologists about *sentiment*, which embraces the entire area of drives, emotions, feelings, and attitudes.

While Homans makes no reference to recent experimental studies of group dynamics, many of his hypotheses are experimentally researchable and should stimulate productive investigation. And the good use he makes of naturalistic observation should remind psychologists of the potential value of nonexperimental approaches in defining profitable hypotheses in this new field.

M. BREWSTER SMITH.

Vassar College.

KUBIE, LAWRENCE S. *Practical and theoretical aspects of psychoanalysis.* New York: International Universities Press, 1950. Pp. xvii+252. $4.00.

In this expanded revision of his book, *Practical Aspects of Psychoanalysis*, first published in 1936, the author addresses himself to prospective patients and to persons concerned with them as well as to physicians and students of psychiatry and psychology. He hopes to explain what psychoanalysis is, clarify misunderstandings, aid in the choice of an analyst, and answer the many questions which are commonly asked about psychoanalysis.

Changes made in this edition reflect alterations in attitudes toward psychoanalysis and the greater sophistication of potential readers which have come about in the fifteen intervening years. There is, as the author says, "less occasion for indignation" because even the misunderstandings which persist are on a more reasonable level. Blind hostility has largely abated. Analysts, moreover, have learned the necessity for variety in theory and practice and for freedom for research. The emphasis has therefore shifted to accommodate the present more informed audience.

Psychoanalysis is defined as "a specific technique for studying and influencing the form, the distribution, and the utilization of psychological forces." Its goal is to broaden the domain of conscious control in human life by uncovering and modifying unconscious psychological forces. In explaining how this is done, considerable exposition of the theoretical structure of psychoanalysis is undertaken. There are discussions of the analysis of the transference, the role of dream analysis, and the technique of free association. In keeping with the psychoanalytic conception of a universal neurotic process, the statistical idea of normality is abandoned. Psychological "normality"—which might better be called "optimal health"—is defined as determination by conscious rather than unconscious forces. Although the necessary theoretical exposition forms an important part of the book, the greatest emphasis is placed upon practical considerations.

Much of the confusion about psychoanalysis arises from the relationship between the analyst and his patient. This is, therefore, discussed at length and in terms which are designed to clear away any suggestion of mystery. The necessity for the "analytic incognito" is made clear, and the transference is explained. The ethical code of the psychoanalysis is stressed. The relationships of the referring physicians, relatives, and friends are gone into thoroughly.

Many readers will be interested in the chapter dealing with cost of psychoanalysis. A psychoanalyst, unlike other medical specialists, must make his entire income from a very limited number of patients and must set his fees according to what these patients can afford. The average fee of $14.50 per session may not produce an unduly high yearly income for the psychoanalyst, but when this fee is multiplied by several weekly sessions over a period of a year, it is higher than the average annual income of a large proportion of families. Clearly, psychoanalysis is an expensive undertaking and one that must be confined, for the most part, to well-heeled citizens. The author is aware of this problem and suggests that the community as a whole is responsible for its solution.

One of the charges commonly leveled against psychoanalysis is that it frequently results in divorce. Dr. Kubie insists that this is an exaggerated impression, and gives cogent reasons. There are also chapters on the contrast between psychoanalysis and faith-healing, and on psychoanalysis and moral responsibility.

In general the author hews close to the line of conservative Freudian theory, deploring such modifications in technique as the presently popular reduction in number of weekly sessions, and such theoretical developments as the emphasis upon cultural factors in the development of neurosis.

In his chapter, "Controversies and Frontiers," he points out the areas in which technical, theoretical, and clinical research are needed. This section will appeal more to his fellow-workers in the field than to the general public, and it should be of interest to students looking for a problem in this field. Professional persons will also be interested in his comments on training matters and on the need for nonmedical psychotherapists.

Dr. Kubie feels that psychoanalysis is the answer to many of the world's most serious problems, and he argues effectively enough for his point of view that he does not seem to be fatuously sponsoring a panacea. He is certainly qualified to define and elucidate his specialty, and the book should be of value to the wide audience to which it is addressed.

Jessie L. Miller and James G. Miller.

University of Chicago.

Fisher, V. E. *The meaning and practice of psychotherapy.* New York: Macmillan, 1950. Pp. xv+411. $5.00.

This is a book of transition, marking a twilight zone between two eras of psychiatric thought. Twenty-five years ago, textbooks on abnormal psychology and psychopathology were still almost entirely *descriptive*—categorical, organically minded, and fatalistic. Fisher's volume points up the contrasting outlook which has come with the newer emphasis upon *psychodynamics* and the possibilities thus opened up for better understanding and more effective treatment.

But the book has a heavy and debilitating imprint of the past upon it. After a brief and fumbling account of what the author considers to be the principles of psychotherapy, which occupies a single chapter of 44 pages, the rest of the book (22 chapters) is devoted to a succession of case histories which are certainly more reminiscent of Kraepelin than of Freud. True, there is throughout a nominal emphasis upon psychotherapy, but it is therapy which is ordered by diagnostic classification rather than by dynamic principles that cut across classifications. For each of the classical syndromes a highly specific type of treatment is described and recommended, as if one were dealing with so many definite and distinct disease entities, for which rational and relatively independent remedies had been discovered.

Many of the clinical portraits are well drawn, and shrewd insights appear on occasion. But in general this book is conceptually so uncoordinated that it will serve as a satisfactory guide to practice only for those who are content to work by rote and prescription rather than within the framework of a consistent, systematic theory of human personality and its vagaries.

O. H. Mowrer.

University of Illinois.

Deming, W. E. *Some theory of sampling.* New York: Wiley, 1950. Pp. xvii+602. $9.00.

Book-publishing in the field of statistics, both theoretical and applied, is undergoing a great and continuing boom. Although there is a certain amount of chaff in this production, in the main the volumes are remarkable for the high level of quality maintained. This high quality level can be illustrated by mention of just a few of the fine publications appearing in the past two years: Cochran and Cox, *Experimental De-*

signs; Mood, *Introduction to the Theory of Statistics*; Wald, *Statistical Decision Functions*; Yates, *Sampling Methods for Censuses and Surveys*. As far as the writer knows, the above-mentioned work of Yates was the first publication in book form dealing rather exclusively with survey *sampling* methods. The book under review is the second; it too continues the high-quality level.

As the title indicates, all the material in this book has to do with sampling. However, treatment of one aspect of sampling—namely survey methods—dwarfs treatment of every other aspect. The survey material might make about 350 pages, a respectable book on its own, and the miscellaneous mathematical topics might make another book—each worth while in its own right.

This work is lengthy and, what with the extensive use of fine type, has even more material than its 600 pages suggest. First, in fine type there are dozens of "Remarks," little helpful asides to the reader telling extra details: just what form of an equation is best for practical work, special cost features of a method, exceptional cases, other examples illustrating use of a method, and so forth. Second, there are many worked-out exercises and problems illustrating in considerable detail the application of the methods discussed. Third, mathematical derivations are shown in great detail with few steps omitted, and frequently alternative derivations of the same result are given, offering the reader that little boost often so helpful. The reviewer hopes this extra care in derivations will not backfire to discourage readers, because such care increases the ratio of equations to text, thereby suggesting need for a higher degree of mathematical sophistication than is actually required for reading the survey material. On the other hand, many texts by slipping in an equation only occasionally and omitting all important steps give an illusion of simplicity of which more wary readers may by now be quite suspicious. To give an idea of the detail—in the discussion of ratio-estimates ten examples of possible uses are given, written out completely, including notation. This thoroughness is emphasized to give an idea of the reference value of the book. The figures look professional, and the equations are as good-looking as a printer can make them. As the author points out, certain topics such as optimum allocation in stratified multistage sampling are barely mentioned. But, on the whole, the coverage is very wide.

After reading this volume, many are going to feel that the good old days of "let's send out a questionnaire," or "let's have a survey" are just about over. The faults of surveys are beginning to be quite well known, and the ways of adjusting these faults are not usually trivial. It begins to appear that an investigator armed merely with a clear conscience and $\sqrt{npq}$ is not adequately equipped for planning, executing, or interpreting factual, opinion, or attitudinal surveys.

Psychologists interested in surveys are bound to be interested in this treatment, which in many ways is more specific and detailed than that of Yates.

FREDERICK MOSTELLER.

Harvard University.

FRYER, D. H., & HENRY, E. R. (Eds.) *Handbook of applied psychology.* (2 vols.) New York: Rinehart, 1950. Pp. xix+382; ix+383–842. $12.50.

In this work the editors have brought together 115 brief essays, averaging about six pages each in length, written by over one hundred contributors and dealing with all aspects of applied psychology. The attempt has been to be all-inclusive, in the sense that articles have been prepared relating to each applied field with which psychologists concern themselves. The attempt has been to be inclusive also in that the editors have tried to provide material on the techniques which are used, the typical patterns of administrative organization, and the desirable training for particular fields, as well as the results of investigation on special topics and problems.

To spread out over this range the book has necessarily had to be spread thin. Thinness is apparent in part in the amount of space allotted to a given topic. Thus, there is a five-page article on "Child Development," a seven-page article on "Aptitude and Intelligence," a six-page article on "Guided Learning," and the like. It has typically been necessary for the treatment of a topic to be either very skeletonized or else very selective, presenting only one or two aspects of the topic. As an illustration of the first, one might mention an article on "Speech Difficulties," where within nine pages of text the author sketches in the characteristics and the therapy for some thirty categories of speech disorder. The second approach is seen in an article on "Product Testing," in which the author limits himself largely to five illustrations of surveys with which he had been personally familiar.

In part, the thinness is apparent in gaps in the areas dealt with. The articles represent shafts sunk here and there, rather than a uniform level of exploitation of the different areas. Thus, in the chapter dealing with techniques of personnel psychology, there is a treatment of validity but no discussion of reliability, an article on item analysis but no consideration of item-writing. There is a section on child development but no treatment of adolescence, adulthood, or old age. Other examples could be cited.

This book cannot, therefore, be considered a fundamental treatise on applied psychology or any section of it. No single topic is treated with a fullness which makes it a definitive treatment, as has been true of some of the past handbooks which have appeared in experimental psychology, social psychology, or child development. In fact, it does not really seem that this was the editors' purpose. The book consists rather of a series of tastes. But who is supposed to be the taster? Is it the beginning student who is taking an introductory survey course in applied psychology? The advanced undergraduate who is trying to decide whether he is interested in psychology as a life occupation? The graduate student who is trying to decide just what form of specialization within psychology he shall follow? It is hard to tell. The variation from article to article is such that the audience seems to be now one, now the other. A three-page discussion of "Drugs and Smoking" would have little value above an introductory course, but not far from this we find several articles which seem to presuppose familiarity with the concepts of factor analysis.

The reviewer was most interested in the latter parts of this work in which an attempt was made to describe the work done by the psychologist in different settings and to outline the training needed for different types of psychological service. Bringing together material of this type seemed to be the most original and distinctive contribution of the document. Some of the articles which gave fairly full discussions of one or two enterprises in which the author had been engaged seemed also to be of interest and value. Articles which provided brief and necessarily incomplete summaries of the research work on a given problem seemed likely to add relatively little to a good text. In general, the reviewer feels that for undergraduate students the book would have been more satisfactory if it had provided more by way of concrete and specific illustrations of the work the applied psychologist does, and less discussion at the level of generalities and of condensed summarization of research work. The beginning student could well profit from more intimate contact with the psychologist at work. The advanced student who wants to be fully informed on a given topic will need to go to more specialized sources than the *Handbook of Applied Psychology*.

Teachers College, Columbia University. ROBERT L. THORNDIKE.

SCHEINFELD, AMRAM. *The new you and heredity*. Philadelphia: Lippincott, 1950. Pp. xxii+616. $5.00.

The first edition of Scheinfeld's *You and Heredity*, published in 1939, won immediate recognition as an outstanding achievement in the writing of popular science. It was selected for distribution to members of the Book-of-the-Month Club and received high praise from scientists in the fields of biology, anthropology, sociology and medicine. The new volume, however, is far better; about 80 per cent of the earlier book has been completely rewritten, and new material to the extent of some two hundred pages has been added. In fact the amount of information here brought together is remarkable, and all of it is presented in such a way as to make it seem important to the reader personally. The result is a degree of reader-appeal that is rarely equalled by a nonfiction book.

One who picks it up for a cursory examination is more than likely to read on and on for hours, and then to lay it down reluctantly.

The book contains forty-eight chapters. The first ten present briefly and with admirable clarity the basic mechanisms of inheritance. Then follow twenty-one chapters which deal with the genetics of eye color, hair color, skin color, body build, multiple births, longevity, physical abnormalities, mental disorders, feeblemindedness, blood types, and specific disease tendencies. Chapter 31, under the title "'Black Gene' Roll-Call," lists more than one hundred fifty kinds of physical and mental abnormalities together with a brief statement of what is or is not known about the genetic factors involved in each. The chapter closes with "forecast tables" for predicting the probable results of matings involving recessive genes, sex-linked recessive genes, dominant genes, and sex-linked dominant genes. This chapter, although lacking the dramatic qualities which characterize so many of the others, is one of the most valuable in the book.

Thus the first thirty-one chapters are devoted almost wholly to the factual results of genetic research. The remaining chapters (which might well have been called Part II) deal with issues that are much more controversial, including, among others, mental tests and their interpretation, special talents, genius, behavior traits, personality and temperament, sexual behavior, criminality, race differences, fertility trends, and eugenics. Because of the wide range of subject matter in these later chapters it was of course impossible for the author to review the available literature thoroughly on any topic. To have done so would have required a volume each for the genetics of racial and class differences, sibling and twin resemblances, special talents and genius, eugenics, etc. It was a choice between treating these controversial issues briefly and summarily or omitting them altogether. In our opinion the author's choice of the former alternative was wise. Precisely because so many of the nature-nurture issues *are* controversial and emotion-provoking, it is all the more important that the general reader should be given some orientation in regard to what is known to be true, what is probably true, what is pretty certainly untrue, and what is still in the realm of speculation.

This is what the author has tried to accomplish, and in the attempt he has maintained a commendable degree of objectivity. Since nature-nurture investigators do not agree among themselves they can be expected to react differently to some of Scheinfeld's interpretations. There are psychologists who will think the author has underestimated race differences in native abilities, but few if any will disagree with his conclusion that even if such inherent differences do exist between race averages, they are overshadowed by the individual differences within a given race. As for IQ differences between the social classes, much the same could be said except that the evidence for genetic factors is here more abundant and less ambiguous. All but the most die-hard environmentalists would agree with Scheinfeld in preferring for adoption the child of gifted rather than dull parentage. We suspect that this preference of the author would have been strengthened by a more critical survey of the literature on adopted children. For example, the two most carefully controlled studies of adopted children—those of Barbara Burks at Stanford and Alice Leahy at Minnesota—are not mentioned; the only evidence referred to in this connection is that provided by the highly questionable studies of Skodak and Skeels at the University of Iowa. However, in his discussion of twin resemblance and of general "educability" the author takes his stand definitely with the moderate hereditarians.

In so wide a coverage of topics, on many of which the data available are far from complete, it would be too much to expect that every tentative conclusion or interpretation of the author will be confirmed by later research. As the author himself points out over and over, genetics is still an infant science. The important thing is that few writers of popular science have ever gone to so much trouble as has Scheinfeld to avoid errors both of fact and of interpretation. He has sought and obtained the help of one or more qualified geneticists on practically every topic treated, especially those dealt with in the first thirty-one chapters. Such wholehearted assistance of the ablest scientists can only be enlisted by an author whose "scientific conscience" is beyond question.

In short, we regard this book as a masterpiece of its kind. If there is a better example of science-for-the-layman we do not know what it is; nor do we know of any book either in the sciences or the humanities that more deserves to be read by every college student, whatever his major field of interest.

Stanford University. LEWIS M. TERMAN.

GULLIKSEN, HAROLD. *Theory of mental tests.* New York: Wiley, 1950. Pp. xix+486. $6.00.

This is an important book—a must for a large number of psychologists. It is not just another book. It is a book that fills a void by bringing together in a unified fashion the scattered literature and new material on test theory. Its author is not only qualified to do an expert job—he did an expert job.

Gulliksen starts off with four chapters on the concept of measurement errors. Fundamental relationships are derived first from a definition of error and then from a definition of true score. Reliability is defined in terms of parallel tests in such a way as to avoid circular reasoning. The exposition on reliability is straightforward, and the reader with only a modicum of elementary statistics will have no trouble following the argument until he encounters, in chapter 5, the explanation of error as interaction. Here a minor slip occurs: whereas the s^2 values prior to page 50 are simple sample variances, the s^2 values in the development on pages 50–57 are unbiased estimates of variance; yet the two types are used interchangeably. Correction for this will lead to minor changes in equations 16, 26, and 39, and in the verbal statements based thereon.

The next three chapters are devoted to the effect of test length on reliability and validity. These chapters are clearly written and include some instructive diagrams. The Thomases who doubt "prophecy" formulas, such as the Spearman-Brown, should read pages 64–67.

In chapters 10–13 consideration is given to the effect of group heterogeneity on reliability and validity, with correction formulas for various types of selection on the ability being tested, on the criterion, and on associated variables. Chapter 13 will be omitted by the reader unfamiliar with matrix algebra, and it is possible that those who can read this chapter, though admiring the neat presentation, may raise the question as to how many ever encounter situations where correction for multivariate selection is feasible. About half of chapter 10 involves an analysis (following Mollenkopf) which purports to show that whether or not measurement errors are correlated with test score depends upon the skewness and kurtosis of the distribution of test scores. The reviewer is highly skeptical of certain parts of this derivation and equally skeptical of the outcome. For instance, the end result simply does not consistently predict the known direction of the relationship between measurement errors and test scores on the 1937 Stanford-Binet.

In chapter 14 one finds the development of a statistical criterion for judging whether tests are parallel, i.e., meet the requirements of equality of means, equality of variances, and equality of intercorrelations (the last in case of three or more forms). The question of the experimental methods of obtaining test reliability is next considered (chap. 15). This is an excellent discussion of the nonstatistical aspect of reliability determination. Next there is a chapter on estimating reliability by way of item homogeneity, and a chapter on speed versus power tests.

Chapter 18, on methods of test scoring (usually "the number of items answered correctly is an eminently satisfactory score"), is followed by a fifty-page chapter on methods of standardizing and equating scores. Here one finds a first-rate discussion of the various scores derived by transformations. The author points out the difficulties encountered in age scales, particularly those difficulties which arise from choice of the regression line used in defining mental age. He does not however, contribute anything as to which regression line is the worse for this purpose. The author has nothing favorable to say for the

MA–IQ scheme of scoring. A more objective account would have listed both the advantages and disadvantages of such scoring.

Chapter 20 is on the problem of weighting and differential prediction. It is first shown that little is gained by weighting items unless the number of items is small. Weighting for optimum prediction, weighting according to reliability, weighting inversely as the standard deviation, weighting by judgment and by factor methods, and weighting to maximize reliability, are all discussed. The dangers involved in using so-called differential tests, leading to patterns or profiles, are pointedly set forth; it is to be hoped that our clinical friends note his discussion.

The final chapter, which is concerned with item analysis, was somewhat of a disappointment to the reviewer, who had hoped to find therein a critique of the many and sundry schemes which are currently advocated and used in item work. Instead, one finds a noteworthy consideration of the important topic of item parameters as related to the mean, variance, reliability, and validity of a test. This sets the stage for the type of item analysis advocated by Gulliksen, and which the reviewer heartily supports. Simply determine for each item its difficulty, its standard deviation, its "reliability index" (defined as its point-biserial correlation with total score multiplied by its *SD*), and its "validity index" (its point-biserial with the criterion multiplied by its *SD*).

Each chapter ends with a neat summary, followed by a set of exercises. A bibliography of some five hundred titles is included. The general typographic style is pleasing, and the text is practically free of the nuisance errors which are so difficult to detect in mathematical material.

Gulliksen has done an excellent, highly commendable job. The few critical points raised by the reviewer are indeed minor in contrast with the innumerable instances deserving praise. This volume represents a valuable contribution; it is a volume of distinction.

Stanford University. QUINN MCNEMAR.

BOOKS AND MATERIALS RECEIVED

ABT, LAWRENCE EDWIN. (Ed.) Human engineering. *Ann. N. Y. Acad. Sci.*, 1951, **51**, 1123–1278.

AMERICAN MISSIONARY ASSOCIATION. RACE RELATIONS DEPARTMENT. *Human rights and human relations. A report of discussions of the Sixth Annual Institute of Race Relations.* Nashville, Tenn.: Fisk Univ., 1951. Pp. 92. $2.00.

BABCOCK, HARRIET. *The MacQuarrie Test as a clinical instrument.* Lancaster, Pa.: Science Press, 1950. Pp. 72. $2.00.

BERGLER, EDMUND. *Neurotic counterfeit-sex.* New York: Grune & Stratton, 1951. Pp. xii+360. $5.50.

BERRIEN, F. K. *Comments and cases on human relations.* New York: Harper, 1951. Pp. xi+500. $4.50.

BLAKE, ROBERT R., AND RAMSEY, GLENN V. *Perception: An approach to personality.* New York: Ronald, 1951. Pp. viii+442. $6.00.

CLARK, WALTER HOUSTON. *The Oxford Group: Its history and significance.* New York: Bookman Associates, 1951. Pp. 268. $3.50.

COBHAM, E. M. *Mary Everest Boole: A memoir with some letters.* Rochford & Essex, Eng.: C. W. Daniel, 1951. Pp. xiv+144. 10*s.* 6*d.*

COLBY, KENNETH MARK. *A primer for psychotherapists.* New York: Ronald, 1951. Pp. viii+167. $3.00.

DEUTSCH, MORTON, AND COLLINS, MARY EVANS. *Interracial housing: A psychological evaluation of a social experiment.* Minneapolis: Univ. of Minnesota Press, 1951. Pp. xv+173. $3.00.

EDUCATIONAL POLICIES COMMISSION. *Moral and spiritual values in the public schools.* Washington, D. C.: National Education Assn. of the U. S., 1951. Pp. x+100. $1.00.

ELLIS, ALBERT. *The folklore of sex.* New York: Charles Boni, 1951. Pp. 313. $5.00.

ELLIS, ROBERT S. *Educational psychology: A problem approach.* New York: Van Nostrand, 1951. Pp. xi+546. $5.00.

FORD, CLELLAN S., AND BEACH, FRANK A. *Patterns of sexual behavior.* New York: Harper & Paul B. Hoeber, 1951. Pp. viii+307. $4.50.

FROEHLICH, CLIFFORD P. *Guidance workers' qualifications: A review of the literature, 1947–1950.* Washington, D. C.: Federal Security Agency, Office of Education, 1951. (Misc. 3376). Pp. 16.

FROMME, ALLAN. *The psychologist looks at sex and marriage.* New York: Prentice-Hall, 1950. Pp. xv+248. $2.25.

GOHEEN, HOWARD W., AND KAVRUCK, SAMUEL. *Selected references*

THIEL, MANFRED. *Versuch einer Ontologie der Persönlichkeit*. Berlin: Springer, 1950. Pp. xxiii+635. DM 29.70.

THURSTONE, THELMA GWINN, AND BYRNE, KATHARINE MANN. *Mental abilities of children*. Chicago: Science Research Associates, 1951. Pp. 48. $.40.

TOLMAN, EDWARD CHACE. *Collected papers in psychology*. Berkeley & Los Angeles: Univ. of California Press, 1951. Pp. xiv+269. $4.50.

TOMAN, WALTER, *Einführung in die moderne Psychologie*. Vienna: Humboldt, 1951. Pp. 268.

TRNKA, B. *A tentative bibliography [of writings on linguistic statistics]*. A publication of the Committee on Linguistic Statistics, of the Permanent International Committee of Linguists. Utrecht-Brussels: Spectrum Publishers, 1950. Pp. 22.

VERNON, PHILIP E. *The structure of human abilities*. New York: Wiley, 1950. Pp. 160. $2.75.

WALLON, HENRI, AND EVART-CHMIELNISKI, EUGÉNIE. *Les mécanismes de la mémoire*. Paris: Presses Universitaires de France, 1951. Pp. viii+116. 400 frs.

WAY, LEWIS. *Adler's place in psychology*. New York: Macmillan, 1950. Pp. 334. $4.50.

WEPMAN, JOSEPH M. *Recovery from aphasia*. New York: Ronald, 1951. Pp. xii+276. $4.50.

WOODRUFF, ASAHEL D. *The psychology of teaching*. (3rd Ed.) New York: Longmans, Green, 1951. Pp. xix+617. $4.75.

WRENN, C. GILBERT. *Student personnel work in college*. New York: Ronald, 1951. Pp. ix+589. $4.75.

YACORZYNSKI, G. K. *Medical psychology: A basis for psychiatry and clinical psychology*. New York: Ronald, 1951. Pp. xviii+535. $6.00.

YOUNG, J. Z. *The life of vertebrates*. London, New York: Oxford Univ. Press, 1950. Pp. xv+767. $8.50.

Vol. 48, No. 6 November, 1951

Psychological Bulletin

EMOTIONS CONCEPTUALIZED AS INTERVENING VARIABLES—WITH SUGGESTIONS TOWARD A THEORY OF FRUSTRATION

JUDSON S. BROWN AND I. E. FARBER
State University of Iowa

The majority of contemporary psychologists would probably agree that the problems commonly associated with the concept of *emotion* must be treated in some detail by any systematic theory of behavior that aspires to even a moderate degree of comprehensiveness. Although a few attempts have been made to delete the term from the vocabulary of psychology, the concept persists in intruding itself into psychological writings, albeit often, following the current fashion, under the headings of conflict, frustration, and anxiety. In spite of the widespread belief in the importance of the phenomena subsumed under these headings and the imperative need for a clear understanding of their influence upon behavior, it is evident that no genuine order can be discerned within this field. Instead, examination of current treatments of emotions reveals a discouraging state of confusion and uncertainty. Substantial advances have been made in recent years with respect to theories of learning and motivation, but the phenomena of emotion have not, as a rule, been considered in these formulations and remain a tangle of unrelated facts. Many reasons could undoubtedly be advanced to account for the amorphous and ill-defined structure of existing theories of emotion; the phenomena are complex, manifold, evanescent, and difficult to measure. It is the opinion of the present writers, however, that the confusion surrounding this concept is primarily the result of serious misconceptions regarding the nature of theory construction in psychology and the kinds of procedures that must be followed if adequate theories of emotion are to be developed and integrated into more comprehensive systems of behavior.

The aims of the present paper are (1) to indicate what appear to be the essential requirements of an adequate theory of emotion (or emotions), (2) to consider the implications of these requirements for certain traditional problems of emotion, (3) to examine briefly some current treatments of emotion in the light of these requirements, and (4) to illustrate the proposed method of conceptualization by presenting the outlines of both a "nonemotional" and an "emotional" theory of behavioral phenomena commonly ascribed to anger or frustration.

I. GENERAL REQUIREMENTS OF A THEORY OF EMOTION

Since there is no unanimity of opinion with respect to what constitutes adequate theorizing in the case of, say, learning, or perception, or motivation, it is scarcely to be expected that the general requirements for a theory of emotion suggested herein will be acceptable to everyone. Without examining the problems of psychological theory construction in detail, therefore, we shall simply attempt to outline the steps to be followed in the light of what appear to be the most adequate techniques available at present. Specifically, we shall attempt to apply to emotion certain of the fundamental ideas of theory construction (2, 15, 16, 38, 39, 40, 41, 42) developed in recent years in connection with more comprehensive theories of behavior.

At the outset, we would insist that the initial step in developing a theory of emotion is the frank recognition that emotion, if it is to enter usefully into scientific thinking, must be regarded as an *invention* or *inference* on the part of the psychologist. Emotion is not a *thing* in the simple, naïve sense that a chair or table is a thing. Like numerous other terms in common use, it cannot be defined by the simple, fundamental operation of "pointing at." It is not, we would insist, a directly denotable empirical datum, as such writers as Köhler (18) have apparently maintained. But to deny to emotion the status of an immediately given simple real is not to imply that the concept is robbed of all scientific meaning or scientific reality. As Boring has observed:

> *Whatever exists as reality for psychology is a product of inductive inference*—usually from experimental data. To say that these realities are hypothetical constructs is not to alter the truth. The atom is a construct and a reality. Its validity is attested by its power of physical subsumption. The realities are always tentative and have to make their way and prove their worth. They are as temporary as all truth. There is no other scientific meaning for reality (3, p. 7).

Thus, we would maintain that emotion is an inference or a scientific construct. Its scientific worth depends upon the degree to which it encompasses relationships among empirical events and thus leads to the more accurate understanding and prediction of behavior.

It has become apparent, furthermore, that at our present stage of theorizing in psychology, considerable progress toward the solution of vexing problems is promised by treating certain constructs as hypothetical variables intervening between environmental conditions and responses. It is in this way, we would suggest, that emotion (or emotions) might most profitably be conceptualized.[1]

The first outline of theoretical procedures explicitly involving hypothetical intervening variables in the formulation of psychological laws was presented by Tolman (42). Subsequently, Hull (16) demonstrated the heuristic value of this conceptual device for the explanation of learning phenomena through the use of such constructs as habit, drive, and response evocation potentiality. Spence (39) has likewise stressed the utility of intervening variables in theory construction and has sometimes referred to such variables as *calculational devices*. They are introduced for the express purpose of helping to account for behavioral phenomena. Their meaning derives solely from the logical, experimental, and computational operations that relate them to antecedent events, to presently acting stimulus conditions, and to responses.

Even if it be granted, however, that emotion, as a term in a science of behavior, should be regarded as a hypothetical intervening variable, the mere postulation of such a term constitutes only the beginning of a meaningful theory. Assuming that the primary concern of a science of psychology is the understanding of behavior at the molar level, then a large part of the psychologist's task is to delineate the factors that function as determinants of behavior. If emotion is to be included as one of the determinants, it is imperative that a certain portion of a theory of emotion consist of hypotheses, or as Spence (40) has termed them, "guesses at laws" concerning the functional relations between the hypothetical emotion and observable behavior. These guessed-at-laws would be of the general form, "behavior is such and such a function of emotion or emotions." Such statements would specify the manner in which different intensities of one or more emotions function as differential determinants of behavior. By means of such functions some of the unexplained variability of behavior can be accounted for in terms of concomitant variability in the hypothetical determining emotional state. This necessary process of working out the relations between emotion and observable behavior is referred to by Hull (16) as the anchoring of the construct to observable events on the consequent side.

But it is not sufficient to limit the treatment of emotion to statements concerning its inferential nature and to specifications of its hypothesized relations to overt behavior. It is also imperative that further specific guesses be hazarded by the theorist as to the manner in which variations in this hypothetical state are related to antecedent and current events and conditions, both external and internal. If such guesses are not made the theory is likely to be circular, in that variations in the state or process that are inferred from variations in behavior may in turn be used merely to account for these same behavioral variations. Tolman, Hull, and Spence have all explicitly and repeatedly emphasized the necessity for anchoring constructs to observable antecedent events. For example, Tolman, referring to Lewin's failure to carry out this important aspect of theorizing, says in part:

---

[1] In the present paper it has not seemed necessary to differentiate between *hypothetical constructs* and *intervening variables* in the manner suggested by MacCorquodale and Meehl (21). For a critical evaluation of their view see Bergmann (1).

> The conceptual system of topology and vectors obviously implies and requires in the end not only a set of principles whereby such and such individual and group life-space configurations will produce such and such behaviors but also *a set of principles whereby the independent variables of environmental situation and given personality make-up will produce such and such inner and outer life spaces.* And it is this latter set of principles which Lewin and his students, while not denying, seem to have been least explicit about stating (43, pp. 3–4. Italics added).

There is yet another requirement, however, that must be met by an acceptable theory. It is essential that the theory contain clear statements of the relationships existing between the construct of emotion and other constructs in the theoretical system. A theory of emotion has limited meaning apart from, and in the absence of, a logical and coherent statement of its relationships to the constructs embodied in more general theories of behavior. One must, we believe, specify in detail the relationship of the variable of emotion to such terms as habit, cognition, motivation, inhibition, or whatever other concepts one may prefer for the explanation of behavior generally. Indeed, it is our contention that the process of fitting emotion into a broader systematic theory of behavior by determining its relations to the constructs of that theory constitutes the essence of what is involved in developing a theory of emotion.

To summarize: a theory of emotion, stripped of its nonessentials, is simply a series of guesses as to the manner in which emotions are presumed to result from environmental events (past and present, internal and external); guesses as to how they may resemble, differ from, and interact with other hypothetical states or processes within the organism; and guesses as to how different emotions, or differing degrees of the same emotion may affect behavior differentially. It is furthermore stipulated that these various guessed-at relationships be stated specifically enough to assure their susceptibility to experimental test.

II. IMPLICATIONS OF THE SUGGESTED THEORETICAL APPROACH FOR SOME TRADITIONAL PROBLEMS OF EMOTION

A number of psychologists have advanced cogent arguments to support the contention that emotion is a superfluous term in psychology. Duffy (7, 8), for instance, has maintained—as have Meyer (26) and Skinner (37)—that emotional behavior differs only in degree from other kinds of behavior. She rejects the concept of emotion because she has been unable to discover any unique, qualitatively different properties of responses that can be unequivocally characterized as "emotional." Though we agree that it is probably impossible to distinguish between so-called emotional and nonemotional responses solely on the basis of their manifest descriptive characteristics, the present approach provides a quite different basis for deciding whether to retain or reject emotion as a distinctive construct. The criterion for making this decision inheres in the suggested requirement that the relations of emotion to other intervening variables in a broader behavioral theory be clearly delineated. *Emotion may be retained as a separate construct if, and only if, it is empirically useful to posit a state or process that is related to antecedent events, to other constructs, or to behavior by a different set of functions from those that characterize other constructs within the theory.* Thus, emotion and some other hypothetical state might have identical properties as behavior determinants and could not, therefore, be differentiated on the response side alone. Nevertheless, it might still be useful to conceive of their differing in respect of their antecedent conditions and (or) in respect of their relations to other constructs. From the present point of view, the problem is not whether emotion does or does not exist. It is a question, rather, of determining whether certain behavioral phenomena can be accounted for satisfactorily in terms of other well-established constructs. If such is not the case, then the introduction of additional "emotional" constructs may be worthwhile.

This line of reasoning is not, of course, restricted to emotion. It applies also to any other terms such as *S-R connection*, *expectancy*, *ego*, *attitude*, *perception*, *consciousness*, or *libido*, which one might wish to introduce into a systematic analysis of behavior. If the psychologist concerns himself with whether or not the ego, for instance, is "really there" in the organism, he is wasting his time. His proper concern is whether by introducing the term ego, by specifying that the ego varies in such and such a manner with observable antecedents, and by assuming that the ego determines behavior in specific ways, behavior can be more satisfactorily explained than if no ego is assumed.

Another implication of the present point of view is that the procedure of categorizing certain responses as emotional and others as nonemotional becomes unnecessary and undesirable. So far as we are aware, previous attempts to achieve such classificatory groupings have been notoriously futile. This is scarcely surprising, since it is hardly to be expected that a knowledge of only the observable features of a response would enable one to make accurate inferences about determining states that, by definition, can be inferred only from a knowledge of antecedent conditions, concomitant states *and* responses. At any given instant, the topography of a given bit of behavior may reflect the action of numerous hypothetical determinants: habits, motivations, inhibitions, response-produced stimulations, perceptions, and (perhaps) emotions. More data than those provided by the topography of a response are needed to enable one to identify the extent of its dependence upon one rather than another of its many determinants. Skinner made this point a number of years ago when he asked: "Why should a certain part of the reactions of an organism be set aside in a special class? Why should we classify weeping in response to a bruised shin as emotional but weeping in response to a cinder in the eye as not?" (37, p. 406). No satisfactory answers to these questions have apparently been suggested.

Psychologists who deal primarily with so-called nonemotional behavior seem less frequently concerned with problems of response categorization than those who deal largely with emotions. In the absence of information about antecedent conditions or environmental events, the learning theorist does not ordinarily ask whether a specific response is due, say, to a habit rather than a drive, nor, usually, whether it is due to one drive rather than another. Were he to do so, he would undoubtedly encounter the same difficulties that have beset students of emotion. For example, if one notes that an animal eats greedily of moist food, one might ask whether the animal is hungry or thirsty. From observations of eating behavior alone, this question could never be answered since this response, like most others, may be evoked, as a result of learning, in the presence of any one of a variety of drives. In the foregoing example, it is even possible that the animal—and this seems to be occasionally true of humans also—is neither hungry nor thirsty but anxious. Information to the effect that the animal had been supplied with food but no water, or the reverse, might enable the observer to make a decision in this instance; but, for the most part, such a discovery would be of much less interest than the determination of the functional relations between the time or kind of deprivation on the one hand and such variables as the rate, latency, or resistance to extinction of eating on the other.

Finally, the traditional problem of determining how many basically different emotions "exist" is from the present viewpoint similar to that of determining whether *any* emotion "exists." The classical approach involved attempts to determine the number of different emotions that could be accurately identified from facial expressions or teased out by painstaking microanalyses of conscious contents. All efforts to reach agreements by these methods have apparently ended in failure. More progress might have been made if it had been asked instead whether observable behavior is most satisfactorily explained by the introduction of one or more constructs carrying an emotional label. If it appears desirable to have more than one, each must be shown to have relations that are in some way distinctive with reference to the other elements of the theory and to each other. Different emotions exist as realities for psychology only if different defining operations and interrelations can be clearly specified for them. If such specifications cannot be made the superfluous terms are scientifically inadmissible and must be coalesced or deleted; appeals for their continuance on intuitive or phenomenological grounds must be rejected. A situation of exactly this sort faces us at the present time in respect to the constructs of conflict and frustration. But a discussion of this problem will be postponed until we have had an opportunity to examine briefly some current conceptions of emotion in the light of the analyses presented above.

III. Some Current Conceptions of Emotions

Since comprehensive surveys of theories of emotion are readily available, no attempt will be made here to present an exhaustive summary or critique of the many divergent views that have been proposed. Instead, a few representative conceptions will be examined in relation to the suggestions for theory construction that have been presented above.

From an inspection of current treatments of emotion, it is apparent that a number of writers have regarded emotion not as a scientific inference but as an existential real. It is somewhat surprising, in view of the numerous excellent papers of the recent past on operationism, that the difficulties accompanying the adoption of such a position require reiteration. The view that "real" emotions exist often includes the notion that they may be only partially encompassed by the specification of all the operations involved in their definition and measurement. Thus the experiential, behavioral, or physiological manifestations in terms of which emotion is typically defined are regarded as but imperfect and blurred reflections of the underlying *ding an sich* (cf. 46). In consequence of this view there have developed the familiar, interminable, unresolvable arguments over questions such as: Which are the "true" or "truest" measures of emotion? How many emotions exist? Can both pleasant and unpleasant emotions be experienced at the same moment? In how many different ways can emotions be expressed or manifested?

As we have already implied, a further shortcoming of several current views of emotion lies in their assumption that certain supposedly distinctive responses or conscious experiences or physiological reactions are uniquely indicative of emotion. To cite but a few examples: according to Shaffer (35) emotions are disorganized responses; for Young (46) emotion is an acute affective disturbance; for Woodworth (45) emotion is a stirred-up state of feeling; for Wenger (44) emotions are certain visceral actions elicited by the autonomic nervous system. As Duffy has indicated, the difficulties involved in attempting to decide from a given bit of behavior, a fleeting (and often unanalyzable) conscious experience or feeling, or a momentary discharge of the autonomic ganglia that an emotion rather than something else is present in a given organism seem almost insurmountable.

The notion that "disorganized" behavior can be unambiguously interpreted as a sign of emotion seems especially difficult to accept. Certainly not all instances of superfluous or ostensibly inefficient behavior can be regarded as signs of emotionality. Were that the case, a great many phenomena now described or explained in terms of such concepts as associative interference, stimulus generalization, spontaneous recovery, hypotheses, and the like would easily qualify as "emotional." Responses mediated by such processes are often ineffective and maladaptive from the view of goal achievement or drive reduction; but most psychologists would agree, we believe, that only confusion would result from their being classified as "disorganized" and therefore as "emotional."

Part of the inadequacy of certain expositions appears to stem from the belief that a simple definition or statement of equivalence, in which emotion is equated to a hypothalamic event, to an experienced feeling, or whatnot, constitutes a *theory* of emotion. Some of the definitions might be defended on the grounds that they are operationally adequate, but this alone would not qualify them as theories in what we would regard as an acceptable meaning of that term. As Koch has so clearly observed: "No matter what degree of precision we attain in the operational definition of our concepts, nothing resembling theory begins to appear until certain functional relationships between the concepts are asserted. Theory appears when these laws or assumptions (usually both) are stated in such a way that testable consequences may be deduced and subjected to experimental verification" (**17**, p. 127). Almost nowhere in the recognized standard "theories" of emotion can one find explicitly stated functional relations of this all-important kind.

The contention that a definition of emotion is a theory might, however, be defensible under the condition that a theory of the *definiens* has already been developed in some detail. Leeper (**19**), for instance, has recently presented a "theory of emotion as motivation." In essence, this proposal involves the rejection of the notion that emotions are disorganizing (or disorganized responses) and the substitution of the view that emotional processes serve as one of the fundamental means of motivation in the higher animals. Though we would agree that it may be useful to attribute motivational properties to postulated emotional states, the unelaborated assertion that emotions are (or are like) motives tells us nothing in the absence of a theory of motivation (the definiens). Since a careful examination of Leeper's paper fails to reveal the details of such a theory, it must be concluded that he has presented at best an incomplete statement of a *theory* of emotion.[2]

But even supposing that a detailed theory of motivation were to be presented, what then? If emotion is regarded as equivalent to motivation in all respects, both constructs must have the same functional relations to observable events and to other constructs. A single term would then suffice for both, and one or the other should be discarded. If emotion and motivation are to be retained as separate constructs it is imperative that *differences* in their relations to at least some other terms of the system be specified.

Although many treatments of emotion are liable to such criticisms as these, there are clear indications that theoretical psychologists are aware of their inadequacies. Tolman, for example, has stated that "In the last analysis each emotion . . . is to be defined not in terms of some obvious and certain 'introspective givens,' but in terms rather of the process's functional character as an immanent determinant [i.e. intervening variable] of behavior . . . " (**41**, p. 268). Hull has implied, though nothing more, that emotions should be treated as intervening variables, considered carefully enough the possibility that response fixation may be explicable in terms of concepts already in systematic use (**9**, **24**). But as a *theory*, it suffers primarily from methodological weaknesses in failing to specify unequivocally the relations between frustration and its antecedent conditions. Thus, Maier found that insoluble problems and punishment did not always lead to frustration and "in some cases fixation also was found to occur under the usual learning conditions . . . " (**22**, pp. 77–78). Clearly, for this writer the criterion of frustration is the appearance of "abnormal fixation," i.e., the very behavior that it was designed to explain. If every instance of fixation is by definition evidence of frustration, and if frustration is regarded as the explanation of the fixation, the "theory" is unassailable; but it is also of little heuristic value.

---

[2] This criticism is not intended to be invidious. A literally complete theory of emotion would, by our criteria, require a very great amount of elaboration. If, as we believe likely, Leeper's theory of motivation would resemble that of Tolman, it might be possible to evaluate his position more adequately by examining the extent to which Tolman's various statements specify the testable relationships that are assumed to obtain between motivational states and observable antecedent and consequent events.

4. As only one example of the extensive treatments of "emotional" phenomena within the framework of Lewinian theory we may cite the work of French (**11**). According to this writer, the presence of opposing forces, one of which is a restraining force, leads to an emotional state termed *frustration*. The strength of this state is assumed to be a function (unspecified) of the strength of the weaker of the two forces when the weaker is greater than a certain minimum. This suggests the way in which an emotional state is related to antecedent events, on the assumption that the relations between the antecedents and the strengths of the competing forces might be specified. As a determinant of behavior, the emotional state is conceived to lead, among other things, to aggression, regression, or to observable symptoms of physical strain and tension. In this instance, however, detailed specifications of the dependency of various kinds of overt responses upon the hypothetical emotional state have apparently not been drawn up.

IV. Nonemotional Interpretations of Frustration Behavior

In order to exemplify the proposed methods of attacking the problem of emotion, the remainder of this paper is devoted to a consideration of phenomena commonly attributed to the "emotion" of frustration. Our principal objective in what follows is the presentation (Section V) of a paradigmatic theory involving the postulation of a special "emotional state" of frustration. But an adequate evaluation of such a theory cannot be made unless some attention is also given to *nonemotional* explanations of the same phenomena. The present section is devoted, therefore, to a brief examination of interpretations of frustration behavior founded primarily upon conventional learning principles. Our selection of frustration for detailed consideration does not involve the assumption that events usually included under this heading are entirely representative of all events to which the term emotion has been applied.

Perhaps the most prevalent assumption about frustration is that it results whenever conditions are such as to hinder or prevent the occurrence of a response. Typically, it is also posited that the responding organism is motivated to make the response, that the response is (or has in the past been) instrumental to the achievement of a goal and that stimuli normally capable of eliciting the blocked response are acting upon the organism. Under conditions of this general sort various types of behavior may be exhibited. Among those frequently reported are: responses of a trial-and-error type directed toward the surmounting of an obstacle, responses of escaping from the frustrating situation, more persistent or vigorous counterparts of the originally blocked responses, aggressive responses, and so-called disorganized responses.

Although responses of these kinds are presumably typical, there is nothing in their topographical features that unambiguously identifies them as "emotional." Descriptively at least, they are much like the reactions that appear in problem-solving or learning situations. This similarity is well exemplified by Dashiell's use of a schematic diagram of adjustive behavior leading to learning (5, p. 364) which is identical in its essential relations with his diagram of frustration resulting from a conflict of motives (5, p. 144). Both schema include factors leading to the blocking of a motivated response, both include persistent activity that is varied and exploratory, and both include barrier-circumventing or "solution" responses and possibilities for readjustment or learning. Melton, following Dashiell, has made a similar analysis of the learning process and has explicitly noted that " . . . this definition and analysis of learning assumes that all instances of learning involve in some degree the frustration of a motivated organism . . . " (25, p. 669).

It would appear, therefore, that there is a high degree of correspondence between frustrating and problem-solving situations, and the reactions observed to occur therein; consequently it might be supposed that explanatory concepts appropriate to the one might also fit the other. Significantly, learning theorists have typically not found it necessary to employ emotional concepts in their interpretations of problem-solving behavior. Rather, they are likely to appeal to the effects of previous experiences (in the same or similar situations), to innate predispositions, to antecedent conditions of deprivation, and the like. To a degree, then, many learning theorists may be regarded as proponents of nonemotional interpretations of frustration.

Additional examples of nonemotional interpretations of behavior in situations where responses are blocked may be found in a number of theoretical analyses of locomotor conflict behavior (4, 14, 20, 27, 34). In these studies, responses of going only part way to a goal, of vacillating, of blocking, or of going out of the field are explained by the same concepts employed in nonconflict situations, i.e. valences, or gradients in the strengths of approach and avoidance tendencies. "Emotional conflict states" are not invoked as *explanatory* devices.

Whether all types of behavior exhibited when reaction tendencies are thwarted can be satisfactorily explained by the nonemotional concepts of the learning theorists is a question that cannot be answered without extensive empirical and theoretical work. Emotional constructs may eventually become necessary; but a strong case could probably be made for the view that current learning theories *do* adequately account for many of the response types listed above as consequences of frustration.

The occurrence of trial-and-error behavior during thwarting has often been interpreted in terms of learning principles. Thus it has been assumed that blocking results in the weakening (extinction) of the dominant response tendency and the appearance of other (initially weaker) members of a response hierarchy. If one of these reactions, e.g. avoidance or circumvention of the obstacle leads to the goal, the tendency to make that response is presumably strengthened. In this way, the organism may learn to "escape" from the frustrating situation. If the strength of the blocked response is very great, however, the organism may persist for considerable periods in its attempts to respond, though in the absence of reinforcement the response may be expected to extinguish eventually.[4]

Increased effort or vigor following thwarting has often been noted by experimentalists. For instance Finch (**10**) has observed that chimpanzees accustomed to pressing the handle of a spigot to get water will press with much more than usual energy if the water supply is shut off. These animals, like human beings, may simply have learned that vigorous responses are likely to be successful under such circumstances. Responses of this sort are also occasionally likely to lead to reinforcement even when the obstacle is another organism, in which case one may be inclined to label the response as "aggressive."

A somewhat different interpretation of these vigorous responses that does not rely upon direct transfer from previous situations might start from the assumption that the animals enter the situation initially with potentialities for exerting widely varying pressures on the spigot handle. Such a response hierarchy could be due either to previous experiences or to innate predispositions. If, at the time the animals are learning to operate the device, responses of pressing the handle lightly are followed by reward, they should be strengthened to a relatively greater degree than equally successful, but more vigorous responses, since the latter involve a greater expenditure of effort. When water is no longer obtained, the gentle responses should tend to become extinguished and to be replaced by the more intense reactions. Thus the behavior of pressing the handle "angrily" can be reasonably interpreted without recourse to any concepts having an emotional flavor, and without necessarily appealing to previous experience in similar situations.

---

[4] If the response was originally learned under conditions of partial reinforcement, it may be particularly resistant to extinction. This phenomenon is accounted for without great difficulty in terms of learning principles. The "abnormal fixations" discussed by Maier can hardly be said to occur in the absence of reinforcement, since they typically involve escape from a blast of air.

A learning interpretation can also be applied to many apparently "senseless," "disorganized" or inappropriate responses that may appear in a frustrating situation. Let it be supposed, for instance, that a specific response has been elicited in a given situation and has been reinforced. If now the subject is placed in a new situation and if there are a number of cues common to the two, this response might be elicited by, or transferred to, the new situation.

Although this explanation might account for the appearance of "frustration" responses in the second of two quite similar situations, it does not explain the occurrence of such reactions in an ostensibly different situation. It might be hypothesized, therefore, that the stimulus elements mediating the transfer of these responses from one situation to a dissimilar one are internal rather than external. That is, the mediating cues might originate in internal events which take place when behavior is blocked, with the same events and hence the same cues being produced by widely differing environmental situations. For example, a rat upon being frustrated for the first time in a bar-pressing situation might "senselessly" bite the bar. This new emotional-appearing response might actually be an old well-established one that had been strongly reinforced during the animal's early experiences with other rats. Certainly in competing for food with other animals a vigorous bite properly administered might occasionally be reinforced by the receipt of food. Although biting a fellow animal and biting a bar differ in terms of the stimulus elements involved, the animal is hungry and anticipatory responses of eating are thwarted in both instances. The thwarting of such responses might afford unique internal stimuli which, through reinforcement, become strongly associated with the response of biting. That responses of biting can be readily conditioned and will transfer to new environments is shown by Miller's experiment

(29) in which rats were trained to bite one another (or a rubber tube) when appropriately reinforced by shock-reduction.

Such analyses as these indicate that nonemotional interpretations may prove adequate to explain a good deal of behavior following blocking. Nevertheless, the possibility remains that a theoretical schema involving the concept of emotion would be more satisfactory. For instance, it may be useful to suppose that thwarting leads to an increase in motivation in order to explain frustration-produced phenomena such as those described clinically as "overcompensation." Such an assumption might also serve as the basis for the notion that escape from a thwarting situation is reinforcing, quite independently of the instrumental function of such escape in reducing some other drive. Finally, the observation that thwarting is accompanied by or leads to rather unique kinds of responses such as aggression, which appear inappropriate to or independent of specific stimulus conditions, might argue for the desirability of postulating an emotional process.[5] In the following section, therefore, a tentative "emotional" theory of frustration behavior is presented as a possible alternative to "nonemotional" interpretations of the type considered here.

V. AN "EMOTIONAL" INTERPRETATION OF FRUSTRATION BEHAVIOR

In the interest of terminological clarity, it should be stressed at the outset that the term *frustration*, as used here, refers to a hypothetical (defined) state or condition of an organism. It is to be distinguished, therefore, from a *frustrating event* such as the blocking or nonrewarding of a response. Unfortunately, the term *frustration* has been used in the past to mean either a state induced by a frustrating event or the frustrating event itself, or both. The terms *anger* or *annoyance* might be substituted for frustration provided they are divested of their classical attributes as unique conscious *quale* or immediately given phenomenological entities.

Since it is our contention that theories of emotion have little meaning apart from more comprehensive theories of behavior, the present treatment has been coordinated with Hull's system as presented in the *Principles of Behavior* (16). Insofar as space permits, the ways in which the present construct of frustration and its relations resemble or differ from those proposed by Hull are indicated in a general manner. Our preference for a formulation of a concept of frustration within the framework of Hullian theory in no sense implies that it could not be treated, perhaps equally well, in terms of other theoretical systems.

---

[5] It should be observed that a demonstration of the utility of adding new constructs or of changing old ones does not invalidate a given theoretical approach in its entirety. One may consider in this regard the manner in which the concept of anxiety, which is strongly connotative of emotionality, has been added to and incorporated within the general framework of Hullian theory.

If the procedures suggested in the earlier sections of this paper are to be followed, the *initial* tasks in outlining a theory of frustration must be (1) the denotation of the antecedent conditions that are assumed to lead to frustration and (2) the specification of the functional relations holding between frustration and these conditions. The first three of the following subsections are devoted to these tasks.

Antecedents to frustration. As we have already noted, a common assumption made by writers in this area is that frustration develops under any condition that hinders or prevents the occurrence of a response to which there is a supraliminal excitatory tendency. With this assumption as a starting point, a number of different kinds of manipulable conditions may be defined as antecedents to frustration. Among these are: (1) the introduction of partial or complete physical barriers; (2) the introduction of delay periods between the initiation and completion of a response sequence; (3) the omission or reduction of a customary reward on one or more trials; and (4) variations in the organism's condition, environment, or training leading to the evocation of a response tendency that is incompatible with an ongoing one.

Although these antecedents differ widely with respect to their manifest topographical features, the capacity they possess in common to thwart an ongoing response may be accounted for by assuming that *each functions in some way to arouse an incompatible reaction tendency.* These competitive tendencies may be either positive excitatory tendencies to perform a conflicting response or inhibitory tendencies such as those resulting from the expenditure of effort. The basic assumption is made, therefore, that *frustration is the consequence of either* (1) *the simultaneous activation of two competing excitatory tendencies or* (2) *the presence of a single excitatory tendency and an opposing inhibitory tendency.*[6]

---

[6] In view of the fact that the oppositional nature of these antecedents is stressed, whether they result from excitatory or inhibitory tendencies, it may be reasonably argued that they should be regarded as *conflict* situations, or in our terminology, antecedents to a hypothetical state of conflict. We do not believe, however, that a useful division can be made at present between conflict and frustration. If one wishes, one might define the state induced by the first kind of antecedent (competing sE_R) described above as conflict, and the second (competing I_R) as frustration. But so far as we are aware, no one has formulated a distinction between the two constructs of frustration and conflict that rests in a convincing way upon demonstrable differences in their functional relations to other constructs or to behavior.

The exact manner in which any particular antecedent might lead to an opposing tendency and the nature of the tendency aroused in each case can only be surmised at present. If an organism has been trained to approach a spatially remote goal and is then hindered by a physical barrier from reaching that goal, it is conceivable that the original approach tendency would be opposed by both excitatory and inhibitory tendencies. For an organism that has had prior experiences in similar thwarting situations the mere sight of the barrier could elicit incompatible excitatory tendencies to retreat. Or, the barrier might lead to an increase in reactive inhibition if it occasioned vigorous and repeated barrier-circumventing reactions. Frustration resulting from the introduction of delay periods is attributable to decreased effectiveness of reinforcement and therefore to smaller increments in excitatory tendency in the face of undiminished increments in work inhibition. Indeed, the increments of inhibition may actually increase, since, in practice, delays are nearly always produced by increasing the distance from the start to the goal or by introducing a physical barrier. In either case, the procedures used to produce delays might provide additional stimuli which could function to elicit or facilitate positive competing tendencies. The way in which nonreinforced trials are conceived to produce frustration depends on one's conception of the nature of the extinction process. Within Hull's theory, reactive inhibition (I_R) is assigned the principal role, though provision is also made for the learning of competing habits ($_SI_R$'s). Finally, all the operations by which an organism's condition, environment, or previous training are manipulated so as to facilitate the evocation of an incompatible response are presumed to lead to frustration mainly through the concurrent activation of competing excitatory tendencies.

The hypothesis that frustration develops under conditions that hinder or prevent the occurrence of a response is a definition of the same general order as Hull's definition of drive in terms of antecedent conditions of deprivation. But unelaborated definitions of this sort have limited applicability. Unless frustration is to be conceived as an all-or-none event, the assumed functional relationship of frustration to its antecedents must be stated more precisely. To do this, both the breadth and the specificity of our assumptions must be increased.

The magnitudes of the excitatory and inhibitory competitive tendencies as functions of their antecedents. Since the assumption has been made that frustration results from the competition of reaction tendencies, and since these tendencies, like frustration, are constructs, it is evident that one of the first tasks is to indicate the ways in which the magnitudes of these hypothetical variables might vary with changes in observable events. Until this has been accomplished, and the further relation of frustration to the competitive tendencies has been outlined, frustration itself cannot be related to manipulable antecedents.

In situations in which the antecedent conditions lead to the elicitation (or strengthening) of a conflicting *excitatory* tendency, the problem of specifying the relations between the observable independent variables and the strength of the competitive tendency is identical with that of accounting for the strength of any response tendency whatsoever. Following Hull, it may be assumed that the competitive excitatory potential ($_sE_R$) is determined by the multiplicative combination of habit strength ($_sH_R$) and effective drive ($\overline{D}$). Consequently the strength of the $_sE_R$ may be assumed to vary in some such manner as Hull suggests with number of reinforced trials, amount of reward, delay of reinforcement, strength of drive, and so on. Since these conditions determine the strengths of both the thwarted and thwarting tendencies they constitute the effective antecedents to frustration when two incompatible responses are instigated.

In situations in which frustration results from the negative effect of *inhibition*,[7] the strength of the inhibition may be assumed to be a function of the degree of work involved in executing the thwarted response and the number of times the response is elicited. Here Hull has proposed as a specific hypothesis that $I_R = cn/B - W$, where n is the number of response evocations, c and B are constants, and W represents work measured in physical units (16). By this equation, if work per trial is held constant, inhibition increases linearly with number of trials; and if the number of trials is held constant, a positively accelerated curve of inhibition is obtained with increases in amount of work. If an additional simplifying assumption is made that amount of work is linearly related to degree of blocking and length of delay, the relation of inhibition to these antecedents and to the number of response evocations might be given to a first approximation by Hull's equation.

Up to this point it has been assumed that the magnitude of frustration is some function of the respective strengths of simultaneously aroused competitive tendencies, and that the strengths of these tendencies may be inferred from Hull's assumptions regarding the specific observable variables influencing the development of excitatory and inhibitory tendencies. The next step is to delineate the functional relations that might obtain between frustration and the competitive tendencies.

Frustration as a joint function of the absolute and relative strengths of the thwarted and competing tendencies. In writings on frustration and conflict two rather frequent, though by no means always explicit, suggestions may be found as to the manner in which frustration or conflict might vary with the strengths of two competing tendencies. It has been assumed (1) that frustration increases as the difference between the strengths of the tendencies is reduced, being maximal at the point of equality, and (2) that if the two tendencies are equally strong, then the greater their absolute strengths, the more intense the frustration. Although there are some experimental data that lend credibility to each of these assumptions separately, the factors involved have not been integrated into a single principle encompassing their joint operation or interaction. For instance, one cannot predict on the basis of the separate assumptions whether two weak tendencies that differ by a given amount will produce the same degree of frustration as two stronger tendencies differing by a greater amount.

---

[7] In this section, the terms *inhibitory tendency* or *inhibition* refer to any state or condition of the organism (other than a competing excitatory tendency) that tends to produce a cessation of a given activity. They correspond roughly to Hull's concept of *reactive inhibition* (I_R). Although Hull's *total inhibitory potential* ($\dot{I}_R$) also includes a conditioned component ($_sI_R$) no attempt is made here to consider the possible influence of this latter factor.

One specific hypothesis with respect to the relation of frustration (*F*) to both the relative and absolute strengths of two competing tendencies may be expressed by the formula

$$F = \frac{E_w^{\,n}}{E_s^{\,n-1}}$$

where $E_w^{\,n}$ is the weaker of the two tendencies (either an $_SE_R$ or an I_R) raised to the n^{th} power, and $E_s^{\,n-1}$ is the stronger tendency (presumably always an $_SE_R$) raised to the $n-1^{\text{th}}$ power.[8] If n is set at unity, F becomes directly proportional to the strength of the weaker tendency alone, since the denominator reduces to unity. It is notable that this specific assumption, though couched in different terms, has already been made by French (11). His proposal thus emerges as a special case of the one suggested here. On his hypothesis, however, neither the absolute strength of the stronger tendency nor the relative strengths of the two tendencies can enter into the determination of frustration. Assuming that such factors as these are important, it is necessary to set n at some value greater than 1 so that F will decrease with departures from equality produced either by increasing E_s (with E_w constant) or by decreasing E_w (with E_s constant).

In Fig. 1 a family of curves has been plotted from the above equation to illustrate graphically the manner in which F is assumed to vary as a joint function of the strengths of the two tendencies. For this purpose n has been set at 2, in which case the equation for frustration reduces to $F=E_w^2/E_s$. In conformity with Hull's procedure, both of the competing tendencies as well as amount of frustration have been placed on a centigrade scale. The weaker tendency (E_w) has been plotted on the abscissa and F on the ordinate, the strength of the stronger tendency (E_s) constituting the parameter.

---

[8] It should be perfectly clear that this particular hypothesis and the specific numerical values employed throughout this presentation have been arbitrarily selected to facilitate the process of exposition. No one would be more surprised than we if these particular hypothetical relations should prove to be exactly the ones that best predict empirical findings in this area. We merely wish to illustrate in an admittedly schematic and incomplete way the kind of quantification of variables and specification of relations that is required, according to our conception, in any serious attempt to formulate an "emotional" (or, indeed, any other) theory of frustration.

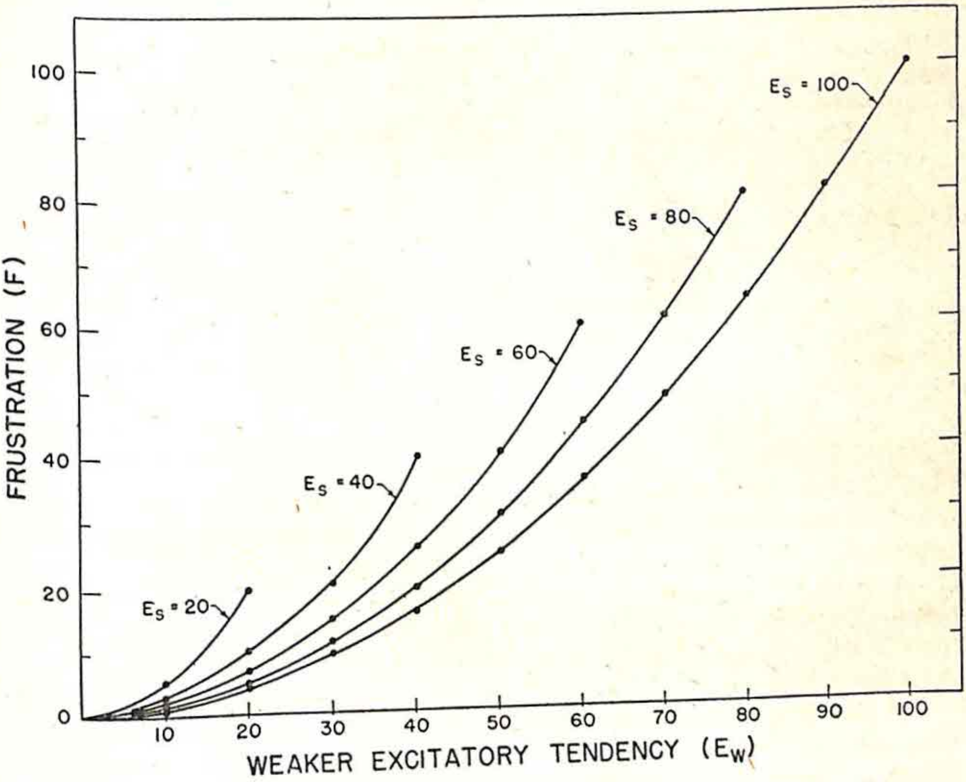

FIG. 1. CURVES INDICATING VARIATIONS IN FRUSTRATION AS A FUNCTION OF THE STRENGTHS OF TWO COMPETING TENDENCIES.

The following relation is assumed: $F=E_w^2/E_s$, where F=frustration, E_w=weaker of two competing tendencies ($_SE_R$ or I_R), and E_s=stronger of two competing tendencies. Each curve represents the assumed growth in the magnitude of frustration with increasing values of the weaker tendency when the stronger is held constant. The successive curves result from the assignment of different values to the stronger tendency.

From these curves, and from a consideration of the equation upon which they are based, the following relations may be noted:

1. An increase in the strength of the competitive tendency (E_w) with the dominant tendency (E_s) held constant, or a decrease in the strength of E_s with E_w fixed, results in a positively accelerated increase in frustration (F) which

reaches its maximum value when $E_s = E_w$. That is, if either of the two tendencies remains unchanged, frustration increases as the difference between them is reduced.

2. Conversely any increase in the difference between the strengths of the two tendencies will result in a decrease in frustration regardless of whether E_s is increased relative to E_w or whether E_w is decreased relative to E_s. These two implications are congruent with the first of the general assumptions about frustration noted at the beginning of this section.

3. Simultaneous increases in the strengths of both tendencies will produce increasingly greater values of frustration provided the increases are such as to maintain either the ratio of the tendencies or their absolute difference constant. In either instance, the relations between frustration and the strength of either of the two tendencies will be linear. This relationship is indicated in part in Fig. 1 by the progressively higher values of F that coincide with the terminal points of the curves for increasing values of E_s. At each of these points the competing tendencies are, of course, equal in strength. This implication relates to the second general assumption mentioned above.

If our initial guesses regarding the relations of the competitive tendencies to their antecedents are now considered in conjunction with the proposed equation for frustration, it becomes possible to formulate a number of specific hypotheses concerning the effect of number of nonreinforced trials, changes in length of delay, degree of blocking, and the like, upon degree of frustration. For example, let it be assumed that I_R increases linearly with successive nonreinforced trials and that the thwarted E_s remains unchanged. It then follows from our equation (by treating I_R as equivalent to E_w) that F will increase in a positively accelerated manner with successive extinction trials. Or again, let it be supposed the E_w increases in a negatively accelerated manner as a function of the number of reinforcements of the habit involved. In this case if E_s remains constant, it can be shown that the curve for the growth of F as a function of successive reinforcements of the weaker habit will have a predominantly negatively accelerated form but with a slight positive initial acceleration. Such hypotheses as these cannot be tested, however, until additional assumptions have been made concerning the relations of frustration to constructs such as habit and drive and, particularly, to overt behavior.

The relation of frustration to the theoretical constructs of habit ($_SH_R$) *and drive* (D). When the competitive tendencies are both excitatory potentials, it follows from Hull's equation for $_SE_R$ (16) that each tendency is determined by its corresponding habit strength ($_SH_R$) and by the effective drive state ($\overline{D}$) which is common to both. Thus,

$$_SE_R = \frac{_SH_R \times \overline{D}}{100}.$$

From this relation, it can be seen that changes in the strengths of either the habits or of the drive will affect the magnitude of frustration. With $\overline{D}$ held constant, an increase in the stronger habit or a decrease in the weaker will decrease frustration; conversely, decreasing the stronger habit or increasing the weaker will increase frustration. If the strengths of *both* habits are varied, frustration may either increase or decrease, depending on the direction and extent of the changes. When the two habits (whether equal or unequal) are held constant, however, frustration will increase with heightened drive and decrease with reduced drive.[9] This holds even though increased drive magnifies the absolute difference between initially unequal excitatory tendencies. Since the numerator is squared in the suggested equation for frustration, any change in drive affects its value more than it does that of the denominator.

Where the weaker of the two opposing tendencies is *inhibitory* rather than *excitatory* the present theory suggests that the effects of motivational changes may be different. If I_R is unaffected by changes in $\overline{D}$, as Hull assumes, and if I_R is substituted for E_w in the equation for F, then raising the value of $\overline{D}$ will heighten E_s only and thereby *reduce* F. Increasing $\overline{D}$ in this situation is therefore functionally equivalent to increasing E_s with E_w held constant. However, it should be noted that an intense drive could lead to more frequent and more vigorous responses of attacking a barrier and hence to the accumulation of more inhibition than would occur under a weak drive. Such an increase in I_R might raise the ratio I_R/E_s to a value equal to or conceivably even greater than that obtaining under a lower drive. In this manner, F might be intensified by increasing $\overline{D}$ even under conditions of competition by an inhibitory state.

Some of the suggested relations between frustration and other constructs may now be summarized. Since frustration is assumed to vary as a function of drive intensity it bears some resemblance to Hull's concept of excitatory potential. The two differ, however, not only in terms of their consequences, but also in terms of their relations on the antecedent side. If drive is held constant, frustration increases with the number of nonreinforced trials, whereas ${}_sE_R$ remains unchanged. (Hull's *effective* excitatory potential, ${}_s\overline{E}_R$, decreases under such circumstances since ${}_s\overline{E}_R = {}_sE_R - \dot{I}_R$). Moreover, although both frustration and inhibition increase with the number of nonreinforced trials, their rates of growth as functions of increases in the value of this variable follow different laws, and their respective effects upon behavior are also assumed to be different.

---

[9] This deduction regarding the expected effect upon frustration of an increase in motivation is supported by Finch's experimental observations (**10**) and by both theoretical and experimental conclusions from the Yale studies of frustration (**6**). Thus, Dollard and his colleagues have assumed that the strength of instigation to aggression (which is roughly equivalent to what is meant here by strength of frustration) varies directly with the strength of instigation to the frustrated response (similar to strength of D in our present usage).

Frustration as a determinant of behavior. The final step in our theoretical outline is to suggest ways in which the postulated frustration state might significantly affect overt behavior. From the hypotheses of other writers and from additional considerations, two likely possibilities emerge: (1) that frustration increases the general level of motivation, and (2) that frustration produces, or is accompanied by, unique internal stimuli.

The possibility that frustration leads to a heightened drive level is clearly coordinate with the notion that emotions have motivational properties, with current conceptions of the drive properties of fear and anxiety, and with numerous, though somewhat unsystematic observations of a clinical nature. Perhaps the most feasible method of incorporating this supposition into Hull's theory would be to assume that frustration-produced drive has *the functional status of an irrelevant drive.* According to Hull, irrelevant drives combine with primary drives to produce the effective drive ($\overline{D}$) that energizes habits into action.[10] On

---

[10] The effect of frustration upon effective drive can be calculated from Hull's formula:

$$\overline{D} = 100 \, \frac{\dot{D} + D}{\dot{D} + 100}$$

where $\overline{D}$ is effective drive, $\dot{D}$ is the irrelevant drive (frustration in this instance) and D, the relevant (primary) drive. If it is assumed that no irrelevant drive is present when the incompatible tendencies leading to frustration are first aroused, then $\dot{D}=O$ and $\overline{D}$ is numerically equal to D. This initial level of effective drive is presumed to determine the strengths of the competing excitatory tendencies and indirectly, therefore, the magnitude of frustration. After frustration develops, however, it is postulated that $\dot{D}$ becomes greater than zero by an amount directly proportional to the strength of F. Now if D is any value less than 100, any increase in $\dot{D}$ as the result of frustration will, by the above equation, lead to a heightened $\overline{D}$.

From the above formula and those presented heretofore, it can be shown that the maximum amount of frustration that can be generated at any fixed level of D can equal but cannot exceed D. This condition obtains when both of the competing habit tendencies are at their maximal strengths of 100. From this relation and from the formula for the combination of relevant and irrelevant drives it can be shown that there is a curvilinear relation between the *increment* added to $\overline{D}$ by frustration-produced $\dot{D}$ and the initial level of D. Thus, although the absolute degree of intensification of reaction tendencies will increase as D is strengthened, *the effective change due to frustration will increase and then decrease* as the strength of the initial dominant drive is increased. This conclusion is consistent with the common-sense notion that when original motivation is weak, frustration is not great enough to affect behavior appreciably, and that when original motivation is strong, frustration, though intense, can add but little more.

this assumption, the presence of frustration should produce an increment in an organism's effective level of motivation. It follows from this that responses elicited at the time frustration is aroused, or shortly thereafter, should be more intense than those elicited when frustration is absent. Moreover, if a response is followed by a reduction in frustration the tendency to perform that response should be strengthened, since drive reduction is assumed to be reinforcing.

Although there are apparently no systematic experimental data to support the contention that frustration may result in the intensification of reactions, fragmentary observations from a number of sources are consistent with this expectation. Already mentioned is the observation by Finch that chimpanzees press on the operating handle of a water spigot with markedly increased vigor when it fails to deliver water. In some of the rate-of-eating curves obtained by Skinner (**37**) there is a clear indication that periods of frustration, produced by temporarily preventing a rat from eating, are followed by an increase in rate of responding. Skinner himself attributes the change in rate in part to an emotional effect, which he relates rather closely to drive. Virginia Sheffield (**36**) has recently adduced evidence indicating that under certain conditions extinction may proceed more rapidly when extinction trials are distributed than when they are massed. In explaining these results she has explicitly assumed that frustration is generated during extinction and that it functions to increase the vigor of whatever response the subject has a tendency to perform. Other evidence that appears to fit the hypothesis of increased motivation due to frustration is found in the observations of agitated or intensified behavior during the extinction of both instrumental and classical conditioned responses (**13**).

It may be observed here that although motivational properties have been attributed to frustration, *it differs from other drives such as hunger and thirst in its presumed dependence upon the arousal of competitive tendencies rather than upon conditions of deprivation. It also differs from other motivational states with respect to the events that result in its diminution.* Frustration, by assumption, is reduced by increasing the strength of the stronger tendency, by decreasing the strength of the weaker, by allowing fatigue effects to dissipate, and so on. Clearly these are not operations commonly supposed to reduce the strength of other drives. In these several additional respects, therefore, the present construct of frustration satisfies the requirement that it have guessed-at relations differing from those obtaining for other constructs already in systematic use.

The second major role that frustration might play in determining behavior is by functioning as a stimulus. If at the time that competing tendencies are aroused, unique, differentiable, internal stimuli are generated, the organism's behavior might be affected in a variety of ways. First, such stimuli could provide the means whereby an organism might learn to discriminate between (react differentially to) frustration and other states such as fear or hunger. Secondly, frustration-produced stimuli might, as the result of innately established connections, elicit specific overt responses or response patterns quite unlike those aroused under other antecedent conditions. Frustration might thus serve as an important *source of new response patterns*. Finally, the attributing of stimulus properties to frustration might provide an acceptable explanation of the fact that so-called "emotional" responses often seem irrelevant and inappropriate to the immediate external situation in which they occur. If a distinctive response were associated, either innately or through learning, with frustration-produced stimuli, it could appear in a wide variety of different situations, providing they all result in frustration. Responses to frustration originally restricted to a single environmental context might thus be elicited by frustration-produced cues in other contexts, despite any ostensible inappropriateness that might characterize such appearances.

Although certain acts might depend initially upon the stimuli accompanying frustration, the mere occurrence of such responses may not always imply the presence of frustration. Like any unconditioned response, a frustration reaction might become conditioned to other (perhaps external) "nonfrustration" stimuli. If so, it could presumably be evoked by these neutral stimuli even in the absence of frustration. This possibility is consistent with our contention that useful inferences as to the presence of frustration must rest upon data other than those obtained from observations of so-called "emotional" reactions alone.

It will be apparent from this treatment of frustration as a determinant of behavior that no attempt has been made to specify precisely which responses will appear in thwarting situations. The present theory does lead to the expectations that the presence of frustration will often result in the intensification of responses, that reactions followed by a reduction in frustration will be reinforced, and that responses associated with frustration in one situation will tend to appear in other thwarting situations. But there is nothing in the theory itself to imply that aggression, for instance, will occur more frequently than withdrawal, or withdrawal more frequently than, say, primitivization. That this is not necessarily a weakness of the theory is suggested by the following considerations.

1. In the first place, an important role is assigned in our treatment to the operation of learning factors. Thus it has been hypothesized that many "nonemotional" responses can become conditioned to frustration-generated, as well as to external, stimuli. And conversely, originally innate reactions to frustration are thought to be capable of becoming functionally connected to almost any stimulus. Starting from such assumptions, it is impossible to anticipate the occurrence of specific responses in the absence of either detailed learning histories for each individual or normative data with respect to cultural determinants of responses and reinforcements. The theory could be altered, of course,

to include an assumption that aggression, or some other type of behavior, is an innately determined consequent of frustration. But since a wide variety of different responses have been observed to accompany frustration it seems highly unlikely, in the case of human behavior at least, that responses to frustration are genetically determined to any important degree.

2. In the second place, an implicit assumption (following Hull) of the present theory is that when two or more response tendencies are simultaneously incited, the one having the greatest momentary strength will be revealed in action, though the vigor, latency, and frequency of the reaction may be modified by the competing tendency. Furthermore, different kinds of competing tendencies may also be expected to affect differentially the qualitative characteristics of reactions to frustration. If a rat is frustrated while running it may respond quite differently from one that is frustrated while pressing a lever. In human subjects, the blocking of gross muscular activity may lead to quite different behavior than does the thwarting of verbal activity. In general, then, we would suppose that actions attending frustration will vary with the nature of the responses evoked by the task or activity in progress. Only from a knowledge of the relative strengths of the response tendencies and their degree of incompatibility could the occurrence of a specific response be predicted.

3. As a final point, it may be observed that the assumed effects upon performance of changes in motivation are such as to impose definite restrictions on our ability to predict frustration reactions. Within the theoretical structure adopted, an increase in motivation will (by virtue of its postulated multiplicative action) increase the absolute difference between two initially unequal excitatory tendencies, providing neither is already at or near its maximum. An increase in motivation due to frustration may lead, therefore, either to better performance, if the "correct" tendency is strongest, or to poorer performance, if the "incorrect" tendency is dominant. We must know in advance what the relative strengths of these tendencies are likely to be if the outcome of frustration is to be predicted in any rigorous fashion. And even this is an over-simplification, since further complexities may result from the elicitation of potentially competitive responses by frustration itself.[11]

---

[11] A further possible complication should be noted. Since frustration is assumed to be a function of the strengths of two (or more) competing tendencies and since these strengths are assumed to be a function of drive strength, then frustration, to the extent that it increases drive, might itself increase the strengths of whatever excitatory tendencies are present, and so lead to ever-increasing frustration. The simplifying assumption is made at present that the magnitude of frustration is uniquely determined by the excitatory strengths of competing tendencies prior to the development of frustration. And although such frustration does increase the excitatory strength of the original response tendencies, it is posited that frustration is not immediately and continuously increased by its own occurrence.

It may be useful at a later time, however, to hypothesize such a "positive feedback," i.e., that one may become frustrated, in part, at least, because one is frustrated. Attempts by the authors to develop a formula specifying the limits or rate of growth of such a self-modifying process have been only partially successful. A more promising attempt to express these relations in a relatively precise mathematical form has recently been made by F. Marzocco. Though the present treatment does not assume continuous self-modification of frustration by itself, we do not believe it to be incompatible with such a notion.

Summary. The major elements of the foregoing tentative theory can be summarized most readily, perhaps, by a diagram such as that in Fig. 2.

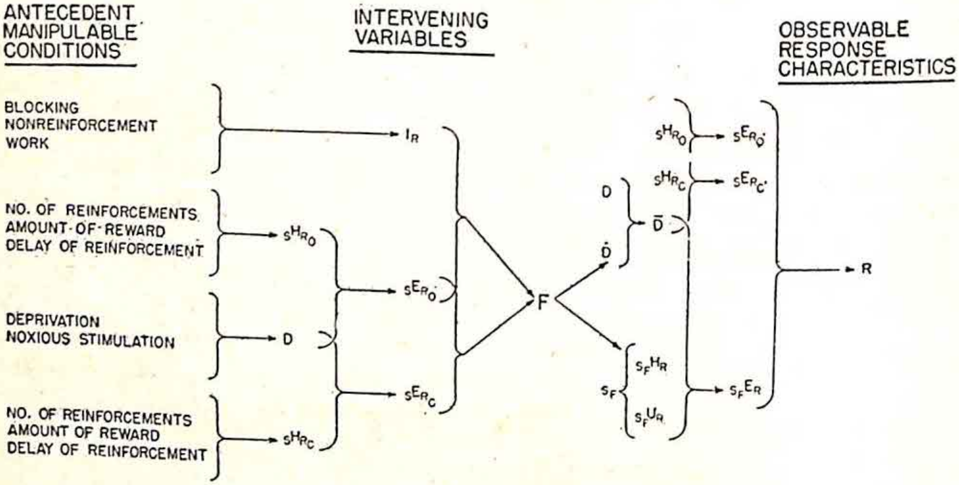

FIG. 2. DIAGRAM SUMMARIZING THE ANTECEDENT MANIPULABLE CONDITIONS AND THE HYPOTHETICAL VARIABLES EMPLOYED IN THE FORMULATION OF THE PRESENT THEORY OF FRUSTRATION.

The symbols for the hypothetical variables are defined as follows: sH_{R_O}=habit strength of an ongoing response tendency; D=relevant drive; sH_{R_C}=habit strength of competing (weaker) response tendency; I_R=competitive inhibitory tendency; sE_{R_O}=initial excitatory strength of ongoing response tendency; sE_{R_C}=initial excitatory strength of competing tendency; F=frustration state; $\dot{D}$=irrelevant drive due to frustration; $\bar{D}$=effective drive strength resulting from combination of D and $\dot{D}$; S_F=frustration-generated stimuli; s_FH_R=acquired habits associated with frustration; s_FU_R=innate response tendencies associated with frustration; $sE_{R_{O'}}$=final excitatory strength of initially dominant response tendency; $sE_{R_{C'}}$=final excitatory strength of initially competing response tendency; s_FE_R=strength of excitatory tendency (either innate or learned) associated with frustration.

At the left of the diagram are shown the antecedent, manipulable conditions that are assumed to lead to frustration through their action as determinants of the strengths of the ongoing excitatory tendency (sE_{R_O}) and the two types of competitive tendency (sE_{R_C} and I_R). The intervening variable of frustration (F) is shown to result from the competition between either sE_{R_O} and I_R or sE_{R_O} and sE_{R_C}. The two major effects of F are indicated as the production of an irrelevant drive ($\dot{D}$) which combines with the relevant drive (D) to produce the effective drive ($\bar{D}$) and the generation of specific frustration stimuli (S_F). These stimuli may have innately associated response tendencies $({}_{s_F}U_R)$ or they may acquire new associative connections $({}_{s_F}H_R)$ through processes of learning. The final consequence of the hypothetical chain of events is the appearance of one or more observable responses, the nature of which will depend upon the relative strengths of the several possible excitatory tendencies.

The adequacy of this particular theory in accounting for the behavioral phenomena so often attributed to the emotion of frustration is an empirical question. It is admittedly incomplete and in some important aspects more obscure than any respectable theory ought to be. Nevertheless, even in this sketchy form, it appears capable of generating a number of deductions that could be put to empirical test. We have no doubt that such tests may indicate the necessity for extensive modification of this theory. Perhaps the facts will even require its abandonment. But, of course, it is precisely this vulnerability to experimental attack that constitutes the end to which our suggested criteria of an acceptable theory of emotion have been proposed.

BIBLIOGRAPHY

1. BERGMANN, G. The logic of psychological concepts. *Phil. Sci.*, 1951, **18**, 93–110.
2. BERGMANN, G., & SPENCE, K. W. Operationism and theory in psychology. *Psychol. Rev.*, 1941, **48**, 1–14.
3. BORING, E. G. *The physical dimensions of consciousness.* New York: Appleton-Century, 1933.
4. BROWN, J. S. Factors determining conflict reactions in difficult discriminations. *J. exp. Psychol.*, 1942, **31**, 272–292.
5. DASHIELL, J. F. *Fundamentals of general psychology.* Cambridge, Mass.: Houghton Mifflin, 1937.
6. DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and aggression.* New Haven: Yale Univ. Press, 1939.
7. DUFFY, ELIZABETH. Emotion: an example of the need for reorientation in psychology. *Psychol. Rev.*, 1934, **41**, 184–198.
8. DUFFY, ELIZABETH. Leeper's 'motivational theory of emotion.' *Psychol. Rev.*, 1948, **55**, 324–328.
9. FARBER, I. E. Response fixation under anxiety and non-anxiety conditions. *J. exp. Psychol.*, 1948, **38**, 111–131.
10. FINCH, G. Chimpanzee frustration responses. *Psychosom. Med.*, 1942, **4**, 233–251.
11. FRENCH, J. R. P., JR. Organized and unorganized groups under fear and frustration. *Authority and frustration: Studies in topological and vector psychology III. Univ. Ia. Stud. Child Welfare*, 1944, **20**, 231–307.
12. HEBB, D. O. Emotion in man and animal: an analysis of the intuitive processes of recognition. *Psychol. Rev.*, 1946, **53**, 88–106.
13. HILGARD, E. R., & MARQUIS, D. G. *Conditioning and learning.* New York: Appleton-Century, 1940.
14. HULL, C. L. The goal-gradient hypothesis applied to some "field-force" problems in the behavior of young children. *Psychol. Rev.*, 1938, **45**, 271–299.

15. HULL, C. L. The problem of intervening variables in molar behavior theory. *Psychol. Rev.*, 1943, **50**, 273–291.
16. HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
17. KOCH, S. The logical character of the motivation concept. II. *Psychol. Rev.*, 1941, **48**, 127–154.
18. KÖHLER, W. *Gestalt psychology.* New York: Liveright, 1929.
19. LEEPER, R. W. A motivational theory of emotion to replace "emotion as disorganized response." *Psychol. Rev.*, 1948, **55**, 5–21.
20. LEWIN, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1935.
21. MACCORQUODALE, K., & MEEHL, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, **55**, 95–107.
22. MAIER, N. R. F. *Frustration: the study of behavior without a goal.* New York: McGraw-Hill, 1949.
23. MASSERMAN, J. H. *Behavior and neurosis.* Chicago: Univ. of Chicago Press, 1943.
24. MCCLELLAND, D. C. Review of N. R. F. Maier, "Frustration: the study of behavior without a goal." *J. abnorm. soc. Psychol.*, 1950, **45**, 564–566.
25. MELTON, A. W. Learning. In W. S. Monroe (Ed.), *Encyclopedia of educational research.* New York: Macmillan, 1941.
26. MEYER, M. F. That whale among the fishes—the theory of emotions. *Psychol. Rev.*, 1933, **40**, 292–300.
27. MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders.* New York: Ronald, 1944.
28. MILLER, N. E. Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. *J. exp. Psychol.*, 1948, **38**, 89–101.
29. MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
30. MILLER, N. E. Learnable drives and rewards. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 435–472.
31. MOWRER, O. H. A stimulus-response analysis of anxiety and its role as a reinforcing agent. *Psychol. Rev.*, 1939, **46**, 553–565.
32. MOWRER, O. H. Anxiety-reduction and learning. *J. exp. Psychol.*, 1940, **27**, 497–516.
33. MOWRER, O. H., & LAMOREAUX, R. R. Fear as an intervening variable in avoidance conditioning. *J. comp. physiol. Psychol.*, 1946, **39**, 29–50.
34. SEARS, R. R., & HOVLAND, C. I. Experiments on motor conflict: II. Determination of mode of resolution by comparative strengths of conflicting responses. *J. exp. Psychol.*, 1941, **28**, 280–286.
35. SHAFFER, L. F. *The psychology of adjustment.* Boston: Houghton Mifflin, 1936.
36. SHEFFIELD, VIRGINIA F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.
37. SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.
38. SPENCE, K. W. The nature of theory construction in contemporary psychology. *Psychol. Rev.*, 1944, **51**, 47–68.
39. SPENCE, K. W. The postulates and methods of 'behaviorism.' *Psychol. Rev.*, 1948, **55**, 67–78.
40. SPENCE, K. W. Theoretical interpretations of learning. In S. S. Stevens

(Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 690–729.

41. Tolman, E. C. *Purposive behavior in animals and men.* New York: Appleton-Century, 1932.
42. Tolman, E. C. Operational behaviorism and current trends in psychology. *Proc. 25th Anniv. Celebration Inaug. Grad. Stud.* Los Angeles: University of Southern California, 1936, 89–103.
43. Tolman, E. C. Kurt Lewin: 1890–1947. *Psychol. Rev.*, 1948, 55, 1–4.
44. Wenger, M. A. Emotion as visceral action: an extension of Lange's theory. In M. L. Reymert (Ed.), *Feelings and emotions.* Mooseheart Symposium. New York: McGraw-Hill, 1950.
45. Woodworth, R. S. *Experimental psychology.* New York: Holt, 1938.
46. Young, P. T. Emotion as disorganized response—a reply to Professor Leeper. *Psychol. Rev.*, 1949, 56, 184–191.

Received February 19, 1951.

A CHI-SQUARE TEST FOR "PRONENESS" IN ACCIDENT DATA[1]

C. J. BURKE
Indiana University

Recently Maritz (2) has shown that certain techniques for the interpretation of accident data can, in some cases, go badly awry. He has constructed examples to illustrate that one can obtain a good "Poisson fit" with accident proneness present in the underlying population and a good "Negative Binomial fit" in the absence of accident proneness. He suggests that proneness be measured in terms of the correlation between two sets of observations on the same individuals.

As Maritz indicates, the correlational technique has one serious drawback. Since the underlying populations are not normal, the usual methods for establishing the significance of a correlation coefficient are inapplicable. Thus, from a given set of data, one is able to report the observed correlation but cannot say whether it could reasonably have occurred by chance in the absence of underlying accident proneness.

Maritz' approach to the problem suggests that the sampling aspects can be attacked by means of the chi-square test. In the present paper, the adequacy of a modification of the chi-square test of independence will be demonstrated. By means of this chi-square test, it can be established whether or not the observed correlation is significant. If it is significant, the correlation coefficient can be used in a descriptive sense as an index of the degree of slope of the regression lines as Maritz advocates.

In the sequel, two cases will be considered. In the first case, it will be assumed that the frequency of accident occurrence in each of the two observational periods is governed by the same Poisson distribution. This would presumably be true if the observational periods were over the same jobs and closely contiguous in time so that the factors causing accidents were relatively constant. In the second case it is assumed that the frequency of occurrence of accidents in each of the two observational periods is governed by a Poisson distribution, but that the underlying probabilities (and hence the parameters of the two Poisson distributions)

---

[1] The author became interested in this problem while serving as a statistical consultant for a research program on accident proneness at Washington University, St. Louis, under contract to the Human Resources Research Laboratory, USAF. He wishes to thank K. S. Teel of the Human Resources Research Laboratory, and W. B. Webb and E. Jones of Washington University for their contributions in a critical discussion of the issues involved.

are different. This would presumably be true if one wished to inquire whether accident proneness in one job situation would carry over into another, if the plant setup were changed between the two observational periods, if the periods were of different lengths or perhaps if the periods were widely separated in time.

CHI-SQUARE TEST—CASE 1

We set up the bivariate distribution of the observed accident frequencies in the two observational periods. Maritz' table V, p. 441, is an example of such a table. We let $F(x, y)$ denote the observed frequency in the cell which corresponds to x accidents during the first observational period and y accidents during the second. We seek to associate with each cell a theoretical frequency $f(x, y; \lambda)$, which has the following properties:

1. The marginal distribution of x is $(\lambda^x/x!)e^{-\lambda}$.
2. The marginal distribution of y is $(\lambda^y/y!)e^{-\lambda}$.
3. x and y are stochastically independent.
4. $\sum_x \sum_y f(x, y; \lambda) = \sum_x \sum_y F(x, y) = N.$ [1]

These properties are satisfied if:

$$f(x, y; \lambda) = \frac{N\lambda^x}{x!} e^{-\lambda} \frac{\lambda^y}{y!} e^{-\lambda} = \frac{N\lambda^{x+y}}{x!y!} e^{-2\lambda}. \quad [2]$$

Then

$$\frac{\partial f(x, y; \lambda)}{\partial \lambda} = f(x, y; \lambda) \left[\frac{x + y - 2}{\lambda}\right]. \quad [3]$$

These derivatives are continuous in λ. Since at least one of these derivatives is different from zero, the rank of their matrix is unity. Then, if we require that none of the theoretical frequencies be too small (always a necessary requirement in a chi-square test), we see that the four conditions of a theorem given by Cramér (1, p. 426) are satisfied. According to this theorem, we may estimate λ by means of the equation:

$$\sum_x \sum_y \frac{[F(x, y) - f(x, y; \lambda)]}{f(x, y; \lambda)} \frac{\partial f(x, y; \lambda)}{\partial \lambda} = 0. \quad [4]$$

In the present instance [4] becomes

$$\lambda = \frac{\bar{x} + \bar{y}}{2} \quad [5]$$

where $\bar{x}$ and $\bar{y}$ are the observed mean number of accidents in each of the samples. When λ has been calculated from the data according to [5], $f(x, y; \lambda)$ can be calculated for all values of x and y which occur according to [2]. When this has been done, a value of

$$\frac{[F(x, y) - f(x, y; \lambda)]^2}{f(x, y; \lambda)}$$

can be computed for each cell and summed over the cells. The quantity obtained in this way

$$\chi^2 = \sum_{\text{all cells}} \frac{[F(x, y) - f(x, y; \lambda)]^2}{f(x, y; \lambda)} \qquad [6]$$

has, according to the theorem cited from Cramér, an approximate chi-square distribution with $(r-2)df$, where r is the number of categories over which the sum extends, whenever the hypothesis of independence is true. (The approximation, as usual, improves as N becomes larger.) If the hypothesis is false, the values obtained would be systematically larger than indicated by the chi-square distribution. Hence, the hypothesis is single-ended and accident-proneness would be indicated by large values of the statistic.

Computationally, one would take as the underlying parameter of a Poisson distribution (λ) the average of the mean number of accidents per individual within each sample. This would, of course, be half of the mean number of accidents per individual in the total sample. With this parameter, two identical Poisson distributions would be established as the marginal distributions in the two dimensional table. The theoretical frequency for each cell is obtained by multiplying the product of the two associated marginal probabilities by the total number of individuals as in equation [2]. The categories in which very small theoretical frequencies occur are grouped so that fairly sizable theoretical frequencies result. An identical grouping is made in the table of observed frequencies. From these (grouped) categories, chi-square is computed, according to [6]. The number of df would be two less than the number of categories.

CHI-SQUARE TEST—CASE 2

We are given a set of observed frequencies $F(x, y)$ similar to those in Case 1 and seek an associated set of theoretical frequencies $f(x, y; \lambda, \mu)$ such that:

1. The marginal distribution of x is $(\lambda^x/x!)e^{-\lambda}$.
2. The marginal distribution of y is $(\mu^y/y!)e^{-\mu}$.
3. x and y are stochastically independent.
4. $\sum_x \sum_y f(x, y; \lambda, \mu) = \sum_x \sum_y F(x, y) = N.$ [7]

The conditions are satisfied if we take:

$$f(x, y; \lambda, \mu) = \frac{N\lambda^x}{x!} e^{-\lambda} \frac{\mu^y}{y!} e^{-\mu}. \qquad [8]$$

Then

$$\frac{\partial f(x, y; \lambda, \mu)}{\partial \lambda} = f(x, y; \lambda, \mu)\left[\frac{x}{\lambda} - 1\right] \qquad [9]$$

and

$$\frac{\partial f(x, y; \lambda, \mu)}{\partial \mu} = f(x, y; \lambda, \mu)\left[\frac{y}{\mu} - 1\right].$$

These derivatives are continuous in λ and μ (as is also the second order mixed derivative) and the rank of their matrix is two. If, as before, we do not permit excessively small frequencies to occur, the four conditions of Cramér's theorem are satisfied. To estimate λ and μ, we have

$$\sum_x \sum_y \frac{[F(x, y) - f(x, y; \mu, \lambda)]}{f(x, y; \lambda, \mu)} \frac{\partial f(x, y; \lambda, \mu)}{\partial \lambda} = 0 \qquad [10]$$

and

$$\sum_x \sum_y \frac{[F(x, y) - f(x, y; \lambda, \mu)]}{f(x, y; \lambda, \mu)} \frac{\partial f(x, y; \lambda, \mu)}{\partial \mu} = 0.$$

After some computation, equations [10] lead to:

$$\lambda = \bar{x}, \qquad \mu = \bar{y}. \qquad [11]$$

When λ and μ have been evaluated from the data by means of [11], a value of $f(x, y; \lambda, \mu)$ can be associated with each cell through equation [8]. When this has been done the basic statistic is computed as before:

$$\chi^2 = \sum_{\text{all cells}} \frac{[F(x, y) - f(x, y; \lambda, \mu)]^2}{f(x, y; \lambda, \mu)}. \qquad [12]$$

If x and y are independent, this quantity has approximately the chi-square distribution with $(r-3)df$ for large N. Again the hypothesis of independence is to be rejected for large values of the statistic.

The computations would be as in the case above except that the two marginal distributions may differ. One would be characterized by a parameter (λ) taken equal to the mean number of accidents per individual in the first sample and the other by a parameter (μ) obtained in the same way from the second sample. The number of df would be three less than the number of categories.

SOME GENERAL CONSIDERATIONS

It is advisable not to group the data before calculating the means. Thus, the estimates of λ in equation [5] and of λ and μ in equation [11] are based on ungrouped data. The data would be grouped to avoid small theoretical frequencies only after the theoretical frequencies have been calculated. In the grouping, one should lump frequencies along the main diagonal of the two-dimensional distribution together as well as frequencies well off the main diagonal. If frequencies near the main diagonal are lumped with frequencies well off the main diagonal, the effect one seeks to detect is masked by the grouping procedure.

One might inquire why the ordinary chi-square test of independence is not used since the underlying distributions within each observation period probably deviate from the Poisson. In this connection, the following points should be noted:

1. If there is no accident proneness the underlying distributions may well be Poisson distributions. Hence, we have followed the usual statistical procedure of basing our test on the conditions assumed by the hypothesis to be tested.

2. Even in the presence of accident proneness the data within an observational period may be well described by the Poisson distribution if the observational period is of short duration. This point is discussed by Maritz.

3. If there are significant deviations from the Poisson distribution in a single observation period this might in itself be presumptive evidence of accident proneness. In this case, we should use the ordinary chi-square test of independence in place of the test developed here whenever deviations from the Poisson distribution are believed to result from changes in the situation.[2] If the deviations are believed to result from individual differences in susceptibility to accidents, we should wish to reject the hypothesis for departures from Poisson distributions in the individual periods as well as for correlations between the periods and the test developed here would be preferred.

AN EXAMPLE

We shall illustrate the computational techniques by means of data gathered by Adelstein and presented by Maritz in his table III (2, p. 438). For convenience, these data are given in Table I of this paper.

We take x as the number of accidents during years 1–5 and y as the number during years 6–11 and use Case 1. From these data,

$$\bar{x}=1.123 \text{ and } \bar{y}=1.123.$$

---

[2] Strictly speaking, however, this test is also inapplicable under these conditions since its applicability is vitiated by variable underlying probabilities. There is perhaps some justification for always using the tests proposed in the present paper. The choice may well be left to the individual investigator. In favor of the test proposed in this paper is the usually larger number of degrees of freedom and the resultant greater power.

TABLE I

Bivariate Accident Distribution Based upon Adelstein's Study

Period 1 (Years 1–5)	*Period 2* (Years 6–11)								*Total*
x \ *y*	0	1	2	3	4	5	6	7	
0	21	14	8	1					44
1	17	12	8	3	1			1	42
2	6	9	2	2	2				21
3	1	1	3	3	1				9
4	1	3							4
5				2					2
6									
7									
Total	46	39	21	11	4			1	122

We take the value of λ equal to the average of these means. Hence

$$\lambda = 1.123$$

The Poisson distribution for this value of the parameter is given in Table II. If two variables are independent, their joint distribution is obtained by multiplying the marginal probabilities. The joint distribution of x and y obtained in this way is given in Table III. Since $N=122$, the

TABLE II

Table of Poisson Distribution

x	*0*	*1*	*2*	*3*	*4*	*5 & larger*
P	.325	.366	.205	.077	.022	.005

theoretical frequencies are obtained from the probabilities in Table III by multiplying by 122. These theoretical frequencies are given in Table IV. The largest frequency $(x=1, y=1)$ has been slightly adjusted to make the sum of the theoretical frequencies equal to 122. Since some of the theoretical frequencies are very small, they have been combined into larger categories as shown by the lines drawn in the table. The lines drawn in Table I represent the corresponding grouping of the observed frequencies. The value of chi-square is computed from the six original

TABLE III

Bivariate Poisson Distribution

x \ *y*	*0*	*1*	*2*	*3*	*4*	*5 & larger*
0	.106	.119	.066	.025	.007	.002
1	.119	.134	.075	.028	.008	.002
2	.066	.075	.042	.016	.005	.001
3	.025	.028	.016	.006	.002	—
4	.007	.008	.005	.002	—	—
5 and larger	.002	.002	.001	—	—	—

categories with fairly large theoretical frequencies and the three categories resulting from the grouping. The data for the χ^2 computation are given in Table V. The probability of this value being exceeded by chance is roughly 0.4. This is not far from the P-value of 0.49 reported by Maritz for a different χ^2-test based on the same data. We should conclude that the correlation of 0.29 reported by Maritz can reasonably be attributed to sampling error. One might suspect, from a category by category examination of the chi-square computation, that the hypothesis of no-proneness could not be maintained if the sample size were increased but this is an empirical matter and cannot be established without further observations. From the given data, there is no reason to reject the hypothesis that the occurrence of these accidents is a purely chance affair.

TABLE IV

Bivariate Distribution of Theoretical Accident Frequencies

x \ *y*	*0*	*1*	*2*	*3*	*4*	*5 & larger*	*Sum*
0	12.9	14.5	8.0	3.0	0.9	0.2	39.5
1	14.5	16.7	9.2	3.4	1.0	0.2	45.0
2	8.0	9.2	5.1	2.0	0.6	0.1	25.0
3	3.0	3.4	2.0	0.7	0.2	—	9.3
4	0.9	1.0	0.6	0.2	—	—	2.7
5 & larger	0.2	0.2	0.1	—	—	—	0.5
Sum	39.5	45.0	25.0	9.3	2.7	0.5	122.0

TABLE V

COMPUTATION OF CHI-SQUARE

f_0	f_t	(f_0-f_t)	$(f_0-f_t)^2$	$\frac{(f_0-f_t)^2}{f_t}$
21	12.9	8.1	65.61	5.1
14	14.5	−0.5	0.25	0.0
17	14.5	2.5	6.25	0.4
12	16.7	−4.7	20.09	1.2
8	8.0	0.0	0.00	0.0
6	8.0	−2.0	4.00	0.5
22	24.2	−2.2	4.84	0.2
11	11.6	−0.6	0.36	0.0
11	11.6	−0.6	0.36	0.0
122	122.0	0.0		7.4

$df=9-2=7$.

SUMMARY

Two statistical tests for the detection of "accident proneness" have been derived. They are chi-square tests which reflect the correlation of accident occurrences in separate time samples involving the same individuals. In both tests a Poisson distribution of accidents is assumed within each observational period. The tests are differentiated on the basis of whether or not the form of the marginal Poisson distribution changes from one period to the other.

It is suggested that these tests be used in conjunction with the correlation coefficient advocated by Maritz in the analysis of accident data. By means of the chi-square tests the significance of observed correlations can be established. The correlation coefficient can then be used as an index of their size.

When significant correlations are observed, some caution should be exercised in locating the reasons for the correlation within the individual. If other factors which may cause accidents are not changed from one test period to the next, the observed correlations may be due to these other factors. As an example we may consider an accident survey covering a large number of workers all operating machines of a given type. Slight differences in the characteristics of individual machines, or differences in their location, may tend to make some of them more dangerous to operate than others. If the operators are not reassigned to individual machines during the experiment, observed correlations may be attributable to the influence of the machines. The significance of the kind of correlation used by Maritz can lead to the conclusion that individual operators are accident-prone only if the experimental design includes random assignment and reassignment of operators.

REFERENCES

1. Cramér, H. *Mathematical methods of statistics.* Princeton: Princeton Univ. Press, 1946.
2. Maritz, J. S. On the validity of inferences drawn from the fitting of Poisson and Negative Binomial distributions to observed accident data. *Psychol. Bull.*, 1950, **47**, 434–443.

Received December 26, 1950.

THREE RATIONAL METHODS FOR THE REDUCTION OF SKEWNESS

RAYMOND H. BURROS[1]
University of Illinois

Psychologists frequently obtain skewed unimodal distributions of scores when quantifying the behavior of organisms in controlled experiments and when gathering norms for tests. It is often desirable to transform such scores in order to get a less skewed, more nearly normal distribution. Three commonly used transformations are the logarithmic, the square root, and the reciprocal.

It sometimes happens, however, that more adequate results are obtained by adding algebraically a constant (positive or negative) to the raw scores before the transformation is applied. Although the best value of this constant can be determined by trial and error, this purely empirical method is tedious and time-consuming. The purpose of this paper, therefore, is to present a rational method for calculating the value of such a constant for each of the three previously mentioned transformations (logarithmic, square root, and reciprocal).

I. *A Logarithmic-Linear Transformation*[2]

Let us suppose that there is a variable x whose population distribution is unimodal and skewed. Let us also assume that there is a variable u which is related to x by means of a logarithmic-linear function:

$$u = \log (x + a), \qquad [1]$$

where a may be positive, zero, or negative. Finally let us postulate that the population distribution of u is unimodal and symmetrical. This implies that

$$Q_2(u) - Q_1(u) = Q_3(u) - Q_2(u), \qquad [2]$$

where the letter Q, with a subscript (1, 2, or 3) means a particular quartile[3] (of u, in this equation).

---

[1] The writer expresses his deepest gratitude to his wife, Helen K. Burros, for preparing the typescript of this paper, which was written at the University of Arkansas.

[2] Although logarithms may be taken to any base, the base 10 is most convenient for computation.

[3] Of course, comparable equations may be written using (in addition to the median) any two quartiles of orders P and 1-P, respectively. For example, the first and ninety-ninth percentiles, or the tenth and ninetieth percentiles, could be selected. The three quartiles are chosen purely for the sake of tradition. Perhaps a more rational basis can be found for a better choice. The reader is invited to consider this problem

Since equation [1] is a monotonic increasing function, any quartile of x is related to the same quartile of u according to equation [1], and we may write

$$Q_j(u) = \log [Q_j(x) + a], \quad (j = 1, 2, 3). \qquad [3]$$

Substituting equation [3] into equation [2] and writing Q_1 for $Q_1(x)$, etc., we obtain

$$\log (Q_2 + a) - \log (Q_1 + a) = \log (Q_3 + a) - \log (Q_2 + a).$$

$$\therefore \quad \log \left(\frac{Q_2 + a}{Q_1 + a}\right) = \log \left(\frac{Q_3 + a}{Q_2 + a}\right).$$

Taking antilogarithms,

$$\frac{Q_2 + a}{Q_1 + a} = \frac{Q_3 + a}{Q_2 + a}.$$

By cross multiplication,

$$(Q_2 + a)^2 = (Q_1 + a)(Q_3 + a).$$

$$\therefore \quad Q_2^2 + 2aQ_2 + a^2 = Q_1Q_3 + a(Q_1 + Q_3) + a^2.$$

Solving for a,

$$(Q_1 - 2Q_2 + Q_3)a = Q_2^2 - Q_1Q_3,$$

$$a = \frac{Q_2^2 - Q_1Q_3}{Q_1 - 2Q_2 + Q_3}. \qquad [4]$$

So far, the reasoning has been done in terms of postulates describing the shapes of the distributions of the infinite population of x and of u, respectively, and the relation between u and x. Suppose now that we have a large random sample of x. We may assume that the sample of x gives a close approximation to the form of the population of x. If we want to reduce the skewness of the large unimodal sample of x, we may calculate the value of a by means of equation [4] and then use it in equation [1] to change each value of x in the large sample to its corresponding value of u. If the distribution of the large sample of u is adequately symmetrical, then the method works. The success of the method can not be guaranteed in advance because equation [2] is a necessary (but not sufficient) condition for symmetry of the population distribution of u. The only way to know whether this method works is to try it. If previous research suggests that a simple logarithmic transformation may help somewhat to reduce skewness, then the logarithmic-linear equation proposed here may well do a better job. If it does not work well, perhaps one of the other methods to be described should be tried out.

II. *A Square-Root-Linear Transformation*

Suppose again that there is a variable x whose population distribution is unimodal and skewed. Let us also assume now that there is a variable v which is related to x by means of a square-root-linear transformation

$$v = \sqrt{x + b}, \tag{5}$$

where b may be positive, zero, or negative. Finally, we shall now postulate that the population distribution of v is unimodal and symmetrical. This implies that

$$Q_2(v) - Q_1(v) = Q_3(v) - Q_2(v). \tag{6}$$

Since equation [5] is a monotonic increasing function, any quartile of x is related to the corresponding quartile of v according to equation [5], and we may write:

$$Q_j(v) = \sqrt{Q_j(x) + b}, \qquad (j = 1, 2, 3). \tag{7}$$

Substituting equation [7] into equation [6], and writing Q_1 for $Q_1(x)$, etc., we find that

$$\sqrt{Q_2 + b} - \sqrt{Q_1 + b} = \sqrt{Q_3 + b} - \sqrt{Q_2 + b},$$
$$2\sqrt{Q_2 + b} = \sqrt{Q_1 + b} + \sqrt{Q_3 + b}.$$

Squaring both sides,

$$4(Q_2 + b) = (Q_1 + b) + 2\sqrt{(Q_1 + b)(Q_3 + b)} + (Q_3 + b).$$

By rearrangement to separate the term with the square root, and by multiplication of terms under the square root sign,

$$2b - (Q_1 - 4Q_2 + Q_3) = 2\sqrt{Q_1Q_3 + b(Q_1 + Q_3) + b^2}.$$

Squaring both sides,

$$4b^2 - 4b(Q_1 - 4Q_2 + Q_3) + (Q_1 - 4Q_2 + Q_3)^2 \\ = 4Q_1Q_3 + 4b(Q_1 + Q_3) + 4b^2.$$

By rearrangement,

$$8(Q_1 - 2Q_2 + Q_3)b = (Q_1 - 4Q_2 + Q_3)^2 - 4Q_1Q_3,$$

$$b = \frac{(Q_1 - 4Q_2 + Q_3)^2 - 4Q_1Q_3}{8(Q_1 - 2Q_2 + Q_3)}. \tag{8}$$

Since equation [6] is a necessary (but not sufficient) condition for the symmetry of the population distribution of v, the method of the square-root-linear transformation can not be guaranteed in advance. As with the previous method, the only way to know whether it works is to try it.

III. *A Reciprocal-Linear Transformation*

As we have done twice before, let us suppose that there is a variable x whose population distribution is unimodal and skewed. Let us also postulate that there exists a variable w which is related to x by means of a reciprocal-linear function:[4]

$$w = \frac{1}{x + c}, \qquad [9]$$

where c may be positive, zero, or negative. Finally, we shall assume that the population distribution of w is unimodal and symmetrical. This implies that

$$Q_2(w) - Q_1(w) = Q_3(w) - Q_2(w). \qquad [10]$$

Since equation [9] is a monotonic *decreasing* function, it follows that

$$Q_j(w) = \frac{1}{Q_{4-j}(x) + c}, \qquad (j = 1, 2, 3). \qquad [11]$$

This means that the first quartile of w corresponds to the third quartile of x and that they are related in the same way by which any two corresponding values of w and x are related, i.e., according to equation [9]. Similarly, the second quartile of w corresponds to the second quartile of x; and the third quartile of w corresponds to the first quartile of x.

Substituting equation [11] into equation [10], therefore, it follows that

$$\frac{1}{Q_2 + c} - \frac{1}{Q_3 + c} = \frac{1}{Q_1 + c} - \frac{1}{Q_2 + c},$$

where we refer to the quartiles of x.

$$\therefore \quad \frac{2}{Q_2 + c} = \frac{1}{Q_1 + c} + \frac{1}{Q_3 + c}.$$

Combining terms on the right side,

---

[4] To avoid the inconvenience of decimal points in the computations, the reciprocal in equation 9 may be multiplied by a suitable power of 10. If such a constant were explicitly introduced into equation 9, it would be cancelled out during the algebraic development and would therefore have no influence on the value of c computed from equation 12.

$$\frac{2}{Q_2 + c} = \frac{(Q_1 + Q_3) + 2c}{(Q_1 + c)(Q_3 + c)}.$$

By cross multiplication,

$$2Q_1Q_3 + 2(Q_1 + Q_3)c + 2c^2 = Q_2(Q_1 + Q_3) + 2cQ_2 + c(Q_1 + Q_3) + 2c^2.$$

By rearrangement,

$$(Q_1 - 2Q_2 + Q_3)c = Q_2(Q_1 + Q_3) - 2Q_1Q_3,$$

$$c = \frac{Q_2(Q_1 + Q_3) - 2Q_1Q_3}{Q_1 - 2Q_2 + Q_3}. \qquad [12]$$

The logic of this transformation is comparable to that of its two predecessors in this paper. Since equation [10] is a necessary (but not sufficient) condition for the symmetry of the population distribution of w, the method of the reciprocal-linear transformation can not be guaranteed in advance. Again, the only way to be sure is to try it.

The Problem of Two or More Samples

The deductions of this paper have been performed in terms of infinite populations. The formulas obtained are therefore applicable only to large samples whose distributions give close approximations to the distributions of the populations from which they are drawn. It is evident therefore that these formulas may not be of much help to the researcher who uses small samples for each subgroup in a factorial design in preparation for an analysis of variance.

Sometimes, however, a research involving two or more large samples is performed. It may happen under these conditions that all the obtained distributions are both unimodal and skewed. If the distributions of these large samples are to be rendered symmetrical without destroying the essential relationships between the samples (for example, if the rank order of the sample medians is to be preserved) then one transformation must be applied to all the measured scores in all the samples drawn. If we assume that one of the three curvilinear-linear transformations presented here will work, then the problem arises of estimating the parameter to be added to each raw score before the chosen curvilinear transformation (logarithmic, square-root, or reciprocal) is applied.

One possibility is to pool all scores into a combined sample and then estimate the parameter from the quartiles of this combined sample. Since the distribution of the combined sample may easily have a different shape from that of any of the separate samples, the value of the parameter estimated from the combined sample may not work well with each sample separately.

Bureau Ednl. Psy. Research
TRAINING COLLEGE
Dated
No.

It appears, therefore, that the only possibility is to compute the estimated value of the parameter for a given transformation from each sample separately. If these values agree fairly well, then perhaps their mean or their median may be used to transform all scores in all large samples of the study. If all the samples are not made sufficiently symmetrical by the transformation used, another of the three discussed in this paper may be tried.

If the estimates of a parameter from several samples differ considerably, then it is doubtful that the chosen transformation is the correct one. Since we know nothing about the sampling distributions of the estimates of the parameters a, b, and c, we can not say how far apart two or more estimates of the same parameter must be before they may be considered significantly different. Further mathematical research is needed on this point.

The Denominators in the Equations for Estimation of the Parameters

By now, the careful reader may have noticed that the denominators in the formulas for estimation of the three parameters, a, b, and c, given respectively in equations [4], [8], and [12], are proportional to the same value:

$$D \equiv Q_1 - 2Q_2 + Q_3, \qquad [13]$$

where the Q's are quartiles of the variable (x) whose distribution is skewed.

This fact presents questions which the writer can not at present answer. Just what are the necessary and sufficient conditions which a curvilinear-linear function, comparable to the three presented here, must have in order to give this result?[5] What other possible denominators are there? Under what conditions will they exist? Why is the denominator the same instead of the numerator? The reader is invited to consider these questions.

SUMMARY

1. Let us assume that a large, skewed unimodal distribution of measurements is obtained in a research. The distribution may sometimes be rendered more nearly symmetrical by first adding algebraically a suitable constant to each raw score and then applying a non-linear transformation function.

---

[5] One possible lead may be suggested. The denominator D is evidently a crude measure of skewness.

2. If the raw scores (x) are transformed by the equation

$$u = \log(x + a), \qquad [1]$$

then the parameter (a) may be estimated by the equation

$$a = \frac{Q_2^2 - Q_1Q_3}{Q_1 - 2Q_2 + Q_3}, \qquad [4]$$

where the Q's are quartiles of x.

3. If the raw scores (x) are transformed by the equation

$$v = \sqrt{x + b}, \qquad [5]$$

then the parameter (b) may be estimated by the equation

$$b = \frac{(Q_1 - 4Q_2 + Q_3)^2 - 4Q_1Q_3}{8(Q_1 - 2Q_2 + Q_3)}. \qquad [8]$$

4. If the raw scores (x) are transformed by the equation

$$w = \frac{1}{x + c}, \qquad [9]$$

then the parameter (c) may be estimated by the equation

$$c = \frac{Q_2(Q_1 + Q_3) - 2Q_1Q_3}{Q_1 - 2Q_2 + Q_3}. \qquad [12]$$

5. Since the derivations of the formulas for estimation of the parameters are all based on necessary (but not sufficient) conditions for symmetry of the distributions of the transformed variables, the success of any of the three methods for reducing skewness can not be guaranteed in advance.

6. The problem of applying any of the three methods when there are two or more large samples is discussed.

7. Several unsolved problems of a mathematical nature are mentioned.

Received March 15, 1951.

STUDIES IN SOCIAL PSYCHOLOGY IN WORLD WAR II[1]

A SPECIAL REVIEW

DANIEL KATZ

University of Michigan

It is difficult for the most enthusiastic supporter of quantitative methods in the social studies to overvalue the contributions of these four volumes to social science in general and to social psychology in particular. The contributions are both profound and numerous. They are essentially of four kinds: (1) the raising of standards for basic and applied research both through the excellence of the research carried on and the full reporting of what was done; (2) substantive contributions to factual knowledge and theory; (3) methodological and technical improvements including both the development of new tools and the ingenious application of old methods to new problems; and (4) advances in the knowledge of the strategy and tactics for conducting social research.

Volume I, *Adjustment during Army Life*, describes the personal adjustment of the soldier as affected by type of experience in the army and by background characteristics. It also treats other morale factors such as job satisfaction, social mobility, attitudes toward leadership and orientation toward the war, with a final chapter of over one hundred pages dealing with attitudes of and toward Negro soldiers. Volume II, *Combat and Its Aftermath*, is concerned with combat motivations, the control of fear in combat, psychoneurotic symptoms, problems of rotation and reconversion, and the point system for redeployment and discharge. It includes a very significant account of the relationship between attitudes toward combat and combat behavior. In Volume III, *Experiments on Mass Communication*, are reported the experiments on the evaluation of the films and educational programs designed to improve knowledge about and attitudes toward the objectives of the war. Methodology is emphasized in this volume both in the discussion of the findings and in a special appendix on measurement problems. Volume IV, *Measurement and Prediction*, is devoted to scaling methods and problems of prediction.

---

[1] Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., Williams, R. M. Jr. *The American Soldier: Adjustment during Army life.* Princeton: Princeton Univ. Press, 1949. Pp. xii+599. $7.50.

Stouffer, S. A., Lumsdaine, A. A., Lumsdaine, M. H., Williams, R. M. Jr., Smith, M. B., Janis, I. L., Star, S. A., Cottrell, L. S., Jr. *The American Soldier: Combat and its aftermath.* Princeton: Princeton Univ. Press, 1949. Pp. 675. $7.50. (Vols. I and II together $13.50.)

Hovland, C. I., Lumsdaine, A. A., Sheffield, F. D. *Experiments on mass communication.* Princeton: Princeton Univ. Press, 1949. Pp. x+345. $5.00.

Stouffer, S. A., Guttman, L. Suchman, E. A., Lazarsfeld, P. F., Star, S. A., Clausen, J. A. *Measurement and prediction.* Princeton: Princeton Univ. Press, 1950. Pp. x+756. $10.00.

The effect of these volumes should be to strengthen the social sciences at one of their weakest points—their general laxness with respect to research standards. The work of the Research Branch forms a marked contrast to much of applied social research in its wholehearted interest in methodological improvement and in the use of adequate scientific communication to achieve this purpose. We are told from time to time of the mines of valuable data in the files of the pollster or market researcher. Seldom, however, does this material see the light of day and, when it does, it is often woefully disappointing in methodological sophistication, in technical reporting and in substantive contribution. Even programmatic research of other social psychologists for government agencies has not had much impact upon social science because it has been very poorly reported to the scientific public. The Stouffer group and the Social Science Research Council in making available the lessons, the findings and methods of this large-scale research undertaking have presented a convincing demonstration that it can be done. It is now incumbent upon scientific workers in social research, even of an applied type, to give back to the field as much as possible of their results and methods. The Social Science Research Council has set a pattern for the foundations not only in stimulating new research but also encouraging the publication and further development of studies already completed. The value of the Research Branch volumes, moreover, would have been increased if they could have had for comparative purposes similar volumes summarizing the research conducted for civilian agencies.

The quality of the scientific reporting is, moreover, high. The conditions under which the data were gathered, the development of the data-collection tools, the statistical devices employed in the analysis of the data, the variance and limitations of the findings, and the construction of new measuring instruments are described in detail. The usual applied survey and even much academic research, not excepting the work of the Warner school, suffers by comparison.

Finally, on the side of presentation, a considerable effort has been effectively expended in maximizing communication through the use of well-constructed tables, ingenious graphical presentations, and unusually well-written text. It is almost as if the authors believed they had someting to communicate and really wanted to communicate it.

The reporting reflects the care and sophistication present in the conduct of the research. In spite of the many practical obstacles to good research in the military organization during the war, a systematic attempt was made to build in checks and controls wherever possible. The logic of the experiment was kept in mind even when it could not be employed. When comparisons were made, samples were matched on relevant characteristics. The studies are more than a straight-run report of the findings but seek to get at relationships by statistical analysis. In fact, the chapter of Lazarsfeld and Kendall in *Continuities in Social Research* (an outgrowth of these four volumes) is the best available treatise on techniques and devices for analyzing survey findings. The net effect of the interest in methodology and in adequate scientific reporting should be to raise the level of scientific standards and aspirations of social scientists both in applied and basic research.

CONTRIBUTIONS TO FACT AND THEORY IN SOCIAL PSYCHOLOGY

On the substantive side the contributions are of three sorts. There is a wealth of descriptive material concerning the feelings, thoughts and perceptions of the large section of the American adult population who constituted the Army. This description of the content of Army culture is, moreover, broken down for different subgroups.

There is also an abundance of findings which illustrate and confirm existing concepts and theories. The contribution here is not in the nature of a new theoretical advance but is rather an indication of the greater generality of known concepts, since the confirmation often comes from new and different settings. Moreover, the data are often analyzed and presented so that the tie-up with theoretical concepts can be readily made. Here, for example, come the demonstrations of the functional nature of beliefs and attitudes; of the importance of primary group standards for the determination and sustaining of beliefs and attitudes; of the significance of psychological reference groups in the determination of judgment; of the influence of role and position in a social structure on the attitudes and behavior of the people playing these roles; and of the cross-pressure phenomenon of conflicting group membership.

Here, too, would be placed a finding related to the Gestalt work on memory which emphasizes not a progressive loss of content according to Jost's law, but progressive transformation of ideas over time. Thus in measuring the effects of orientation films on the opinions of troops, there were in some of the measures greater effects several weeks after the film than several days after the presentation. The limited effects of information communicated through mass media, no matter how well presented, are clearly demonstrated. Information increased specific knowledge but was not generalized to attitudes or to motivation about the war. This is a striking parallel to the studies of OWI on the information campaigns on the home front. These campaigns increased specific knowledge but generally had little effect on basic understanding of fundamental issues or on behavior other than that specifically suggested by the campaign.

In the third place, there are new findings which have a great deal of relevance for the reformulation of older theories or the selection of appropriate theories where there is a conflict in theory. These studies, for example, contain the best evidence we have on changing the attitudes of whites toward Negroes through environmental manipulation. Theories of prejudice have tended to emphasize the pictures in men's heads which bear no relation to objective experience. The older point of view emphasized stereotypes; the newer approach stresses perceptions arising from depth factors in the personality. The Research Branch studies show, however, the direct relevance of objectively based experience in the modification of attitudes. The greater the degree of actual contact with Negroes under conditions of combat, the greater the favorable attitude shown by white soldiers toward them. In other words, any theory of the development and modification of prejudiced attitudes must pay more attention to the world of objective realities in which experiences are formed. If the white child grows up in an environment in which Negroes are assigned inferior roles, menial tasks, slum housing, etc., his prejudice may be related to these experiences. To change his attitude it may be just as necessary to change existing social realities as it is to psychoanalyze him.

Suggestive findings for the development of theory of attitude formation and change come too from the experiments on the effects of mass communication. The presentation of both sides of an argument was found to produce more change than the presentation of one side for the men whose initial position was opposed to the one-sided presentation. It was easier to persuade men to change their minds, if their own point of view was espoused, as well as the opposing point of view. Those who believe in the importance of cognitive structure can interpret this finding as evidence for the principle of imbedding in an accepted context new alien materials to produce restructuring. In fact, the whole series of experiments on mass communication argue for a program of research on attitude change which will take more adequate account of such characteristics of attitude as cognitive structure, crystallization, motivational function, and telic dimension.

Other examples of findings of far-reaching significance for social-psychological theory are the results on the relationship between attitudes and overt behavior. Modern psychological theory makes attitudes the intervening variables on the basis of which specific actions can be predicted. Objections have been raised to this theory on the ground that verbalizations of attitudes represent rationalizations which bear little relation to what the individual will do. The problem is further complicated by the methodological difficulty of equating an attitude expressed in a questionnaire or an interview with an attitude expressed in a natural social situation. There has been little systematic work on the basic issue of the relationship between the verbalization of the attitude and the behavior relevant to it. The Research Branch studied soldiers' attitudes toward combat, however, in relation to the non-battle casualties of companies in combat. The companies scoring highest on questions about willingness for combat, confidence in combat stamina, and confidence in combat skill had the lowest non-combat casualty rates. Even though the correlations were on a group basis and ranged in average from −.27 to −.48, the findings are striking when one considers the crudeness of the criterion measure and the many reasons why men might not report accurately on the attitudinal questions. The theoretical assumption that attitudes, as measured by verbal responses to a questionnaire, are general predispositions to behave in a given manner receives unexpected confirmation from the studies of the American soldier.

To indicate the relevance of the research results for social-psychological theory is not to assert that the objectives of the studies were always formulated in the soundest theoretical fashion. The authors, for example, distinguish between morale as a group concept and personal adjustment as an individual concept—an important distinction. Morale is conceived of as cooperative effort toward some common goal, but the authors forget that such cooperative effort reduces to the effort of individuals and hence is a function of the individual's acceptance of the group goal, his acceptance of group pathways toward the group goal and his related motivations to achieve the common objectives. Thus when measurements are made of personal adjustment to the Army, three of the four areas investigated are personal commitment to the war effort (willingness to serve in the Army and to make sacrifices), satisfaction with job and status, and criticism or approval of the Army. Many of the questions asked in these areas of "personal adjustment" clearly involved attitudes of acceptance of the group goal and of common pathways for reaching it. In fact, in Volume III some of these very questions of personal adjustment are used as measures of morale attitudes for the combat flying personnel of the Air Corps. The inadequate conceptualization of morale means that morale is treated only as a dependent variable and not as an intervening variable on the basis of which to predict behavior.

METHODOLOGICAL CONTRIBUTIONS

The methodological progress achieved in the studies of the Research Branch is to be found in the third and fourth volumes, dealing with experiments on mass communication and with refinements of scaling procedures.

Research design. In an experimental evaluation of army films a number of experimental designs were employed. In addition to the conventional before-after design which may affect the group through alerting them to the purpose of the experiment and predisposing them to a consistency of response, an after-design was employed. Half of the group was exposed to the film and then measures were taken of the whole group. Though weaker in some respects than the before-after pattern, this procedure has the advantage of not affecting the experimental group through the initial measurement. In addition to the practical economy of the one measurement a very limited methodological comparison of the two designs gave some indication of a greater effect in the after-design because of less consistency of response.

Scaling. The major methodological contributions are in the field of scaling. The pioneer work of L. L. Thurstone in developing the logic, the mathematics and the techniques for the measurement of qualitative data some twenty years ago was perhaps premature in relation to the development of the field. Since the work of Thurstone, social psychologists and social scientists have tended to deny, ignore or postpone a systematic consideration of these problems. The Research Branch, however, has boldly given this area of measurement the central attention it deserves. Two models are presented in some detail—the scalogram analysis of Guttman and the latent structure analysis of Lazarsfeld. In his introductory chapter to the volume Stouffer takes the statesmanlike position that the intent is to accelerate the development of this methodological area and that revisions of the models presented—or, indeed, completely new models—are needed.

In addition to the advantage of the scalogram technique for obtaining a unidimensional scale of the sign character of an attitude, Guttman and Suchman present a device for the dimension of intensity and use it to give a zero point for the content scale, thus making possible the comparison of groups on different questions. The zero point is reached as one moves down the content scale with decrease in intensity until the point where intensity begins to increase again. Guttman hopes, moreover, to find additional dimensions of attitudes through scalogram procedures. The scalogram analysis of Guttman has been criticized for its limited practical uses. Reproducibility of an order to give a scale or a quasi-scale is not easy to obtain save for fairly repetitious questions. A more fundamental criticism of the Guttman method is its empirical and mathematical character. Though useful as a technique one hesitates to endow it with as much theoretical importance as does its author. It does not represent a social-science or psychological model, in that there is no theoretical scheme for the interrelationship of variables. The dimensions of attitude are to be derived operationally without reference to psychological theory. Thus the attitudinal variable about which Guttman seeks unidimensionality is defined merely as a delimited area of behavior. What the "more or less" signifies on the Guttman scale in terms of psychological processes is conspicuously by-passed. Thus when the intensity function is suggested as a second dimension of attitudes we do not know to what extent this is a repetition of the first dimension. More empirical tests are needed to validate the independence of the content and intensity components.

The summary dismissal of Thurstone's contributions in Guttman's presentation seems unwarranted. Thurstone did more than develop scales on the basis of the evaluation of judges. He also addressed himself to the problem of the ordering of people on the basis of their own reactions, as in his rank-order and paired-comparison procedures and especially in his method of similar attributes.

The Lazarsfeld latent structure analysis is another procedure for discovering the components of attitudes through mathematical means. Lazarsfeld does not dismiss the illogical opinions of people on a problem as error from the point of view of scaling but assumes a more complex pattern of underlying attributes than the unidimensional scale. To uncover these latent attributes he employs the logic of factor analysis but not its correlational techniques. In its present stage of development, however, latent structure analysis is still a laborious method.

The value of these scaling procedures is two-fold: (1) they contribute methods of determining unidimensionality of the attribute being measured. Thus they can be used in their simpler forms to improve measures where we seek interrelationships between variables defined by existing psychological knowledge and theory; (2) like factor analysis they can suggest basic factors responsible for observed relationships and thus lead to hypotheses for experimentation. They are basically inductive, however, and seek to discover variables or components through their differential mathematical effects.

STRATEGY OF SOCIAL RESEARCH

Perhaps the real significance of these volumes is the indication they give of one direction in which social science can make substantial progress in the near future—namely, pushing for methodological improvement and for theory building within the framework of research concerned with practical problems. Since the Research Branch studies represent a high level of applied research, they have been widely acclaimed for their systematic empirical findings and for their application of rigorous quantitative methods to important social problems. But because they have been essentially *applied* rather than *basic* in character they have been caustically criticized for their empiricism and for dressing up as social science a limited type of social engineering. Critics who believe that scientific progress comes from the testing of hypotheses in the interests of theory construction and reformulation are skeptical of the limitations of research designed to produce specific truths for some Army commander.

With the objectives of these critics, the development of general laws and principles, there can be no quarrel. The real problem consists of the strategy and tactics to achieve this goal. Social science needs the support and approval of its experimental subjects in a fashion not true of the natural sciences. Even if we had huge funds to conduct field experiments in the interests of pure theory in social psychology, we would run up against the limitations of the cooperation of our subjects. We cannot produce experimental changes in a community or in an industrial company or in an army without gearing into on-going programs of changes already acceptable to many. The more we can demonstrate the value of our methods to the nonacademic world in relation to their problems, the more freedom we can win for our own theoretical work. The Research Branch demonstrated the value of social science methods in convincing fashion in their solution of redeployment and discharge through a point system based upon empirical research. Demonstrations of this sort are needed to give status, support, and autonomy to the social scientist.

Another reason for emphasizing the importance of direct contact with practical problems is that such problems are closer to our subject matter than are the practical problems of engineering to the physicist. In the field of personality the most fruitful sources of hypotheses still derive from Freudian theory which had its origin in the study of important problems of human adjustment. Then, too, we are likely to find that more significant manipulation of social and psychological variables occurs in the practical events outside our control than in any manipulations we ourselves can or want to devise.

The strategy of social research should, therefore, not be a complete concentration upon pure science and a postponement of all effort until we have a full-spun theoretical framework and a completely free hand in some ideal community created for experimental purposes. Nor need we declare a moratorium on laboratory experimentation. The interaction of basic and applied research should prove more productive than the concentration upon the one or the other with resulting isolation of theory from practice. And, as the Research Branch has shown, we can make progress in the immediate future by utilizing the resources so readily available for applied social research in as broad a scientific framework as the development of the field permits.

Received July 26, 1951

LEWIN, K. *Field theory in social science: Selected theoretical papers.* (Edited by D. Cartwright.) New York: Harper, 1951. Pp. xv+346. $5.00.

This volume, together with the previously published *Resolving Social Conflicts*, is intended to bring together for convenient reading a number of Kurt Lewin's papers published while he lived in the United States. Devoted to his major theoretical papers of the period, it is more technical than the previous collection. Cartwright has taken his editorial responsibilities seriously, arriving at a surprisingly coherent presentation that provides the best available statement of Lewin's final systematic position. As Cartwright writes, "for the most part it [was] possible to leave the single papers without editorial modification, [although] certain duplication had to be omitted here and there and certain transitional paragraphs . . . supplied . . . ." Three of the chapters involved more substantial rearrangement. Cartwright has added a helpful interpretative foreword.

The first three chapters set the stage by considering important issues in the philosophy of science from the perspective of field theory. The six chapters that follow apply this perspective in selected research areas: learning, development and regression, social psychology, "psychological ecology," and group dynamics. The chapter on "Behavior and Development as a Function of the Total Situation" from the *Manual of Child Psychology* provides a final distillation of the Lewinian findings.

Only four years after Lewin's untimely death, it is still much too soon to take the measure of his contribution to the development of psychology and the social sciences. By bringing scattered papers into a coherent focus, the present volume at once serves to keep us in the presence of his catalytic originality, and contributes to the understanding and eventual appraisal of his distinctive views.

For one thing, it is clear that the "field theory" that provides the common theme of these papers is no theory in the sense of a more or less tightly formulated conceptual system.

> Field theory . . . can hardly be called correct or incorrect in the same way as a theory in the usual sense of the term. *Field theory is probably best characterized as a method:* namely, a method *of analyzing causal relations and of building scientific constructs* (p. 45, italics Lewin's).

While Lewin variously lists the criteria of field theory (for instance, see pp. 60–64) his distinctive approach is best seen as he examines now a traditional psychological concept, now an unconventional research problem, always seeking conceptual clarity that does psychological justice to the causal or functional relations inherent in the phenomena in question.

The graduate student who encounters Lewin through this book can avoid several fruitless controversies occasioned by his earlier volumes. The issue of contemporaneous vs. historical causation turns out, it seems to me, to have been for the most part a red herring; likewise the issue of statistics in psychology. Topological or hodological representation, moreover, no longer appears as the *central* feature of Lewinian theory. It rather stands as his most ambitious attempt at constructing a formal language to deal with problems that he was always ready to approach in other ways—more *ad hoc* or more promising according to the reader's tastes. The student will have to turn to Lewin's other volumes for extended treatment of Lewinian topology.

The question of whether Lewin's mathematical baggage earns its freight will remain with the reader of this book. I share a common student reaction in feeling that it only clutters an otherwise lucid passage, for example, to translate the statement that "the variety of behavior increases during childhood with normal development" into the "formula"

$$\text{var}\,(B^{Ch}) < \text{var}\,(B^{Ad}) \quad \text{(p. 100)}, \tag{1}$$

—and there is much of this. The provocative and valuable analyses in the book do not tax the resources of good English, while the flights of purely formal derivation, particularly the treatment of unity, dependence, and differentiation in the Appendix, seem sterile in comparison.

The consecutive reading of these papers left me with a final reaction: that Lewin's firmest research contributions may lie not in his conception of the total life space as a basis for exact prediction, but in his application of the field approach to much more restricted problems. As it has often been noted, Lewin never managed to describe a *total* life space independently of the behavior to be predicted; there are examples in this book of *post hoc* interpretations. On the other hand, the careful analysis of Zeigarnik's research that occupies most of Chapter 1 shows the fruitfulness of his approach in posing and illuminating a significant cluster of problems in connection with which neither the total personality nor the total situation needs to be described. In spite of Lewin's occasional protestations, it may well turn out that the modest clarification of concepts and demonstration of formal and empirical relations among them is in the long run a far more important contribution than the programmatic objective of exact prediction from the total situation. Values of this sort, at any rate, are amply realized in the present book.

M. BREWSTER SMITH.

Vassar College.

FESTINGER, L. S., SCHACTER, S., & BACK, K. *Social pressures in informal groups.* New York: Harper, 1950. Pp. x+240. $3.00.

This volume presents results of "a study of human factors in housing" from the Research Center for Group Dynamics, University of Michigan.

Part I reports in nine chapters the research study of a housing community for married veteran students at Massachusetts Institute of Technology, for the purpose of determining the effect of housing arrangements upon friendship patterns, group formation, structure, and standards, and the process of communication. It concludes with a theoretical description of the process of development of group structure and standards.

Part II contains two interesting and informative chapters, each of which was written by a practitioner in the field of housing describing the implications of the research for housing planning and practices.

An appendix presents the researchers' rationale of the methods used in the field studies, *viz.*, informants, participant and non-participant observers, informal and standardized interviews, sociometry, and field experimentation. Psychologists will want to read this section before going very far into the book.

Sociometric choices (social friendships) of residents were systematically studied in reference to ecological factors. Physical distance between homes, design and structural factors such as location of face of house or staircase, corner houses, and similar features were shown to influence the opportunities for contact between residents and subsequent friendships.

The problem of adequately describing patterns of group structure and clique formation was partially solved by the development of an interesting technique which should have considerable sociometric usage. It involved the squaring and cubing of matrix-presented sociometric patterns to reveal indirect connections between persons (e.g., influences which are transmitted from friends of friends) and the existence of cliques.

The theoretical statement of the ways in which group structures develop concerns the importance of the type of initial person-to-person contact, physical surroundings, friendship patterns, and length of residence. Group standards are developed primarily from the friendship patterns and the resulting communication patterns which tend to produce a high degree of uniformity in the information.

Throughout this book the pattern is the same: theoretical deductions are made from general hypotheses and the data are re-examined to verify or reject. The writers are skillful theorists.

However, this reviewer could not shake off some nagging questions which never seemed to get answered. One such question concerns how the original contacts between two residents were made. The underlying assumption used for the entire study was that there was a constant probability over the entire project for a contact to lead to the establishment of a friendship (p. 153). In addition, ecological factors are supposed to be the determiners of these contacts (p. 58). A constant question was whether graduate or undergraduate standing, or departmental affiliation, had any effect at all on the friendship patterns which were so basic to the whole study.

Another question had to do with turnover among the residents. Were the persons who deviated from the attitudes and standards maintained by the court in which they lived a function of the length of residence in that court? Not only the analyses of the deviates but the friendship patterns seemed to depend upon an answer to this question.

It is hoped that the investigators will continue to study face-to-face groups to develop our insights into such critical areas of human behavior. Their results thus far are quite promising and in this study already show signs of influencing the views of housing practitioners.

RALPH R. CANTER, JR.

University of California.

DEUTSCH, MORTON, & COLLINS, MARY EVANS. *Interracial housing: A psychological evaluation of a social experiment.* Minneapolis: Univ. of Minnesota Press, 1951. Pp. xv+173. $3.00.

This book reports the first of a program of studies in intergroup relations being conducted by the Research Center for Human Relations at New York University. It is an impressive piece of research that will contribute significantly to the growing stockpile of knowledge about intergroup relations, as well as provide a sound factual basis for guiding official policy decisions relating to the occupancy pattern of new and existing public projects. In fact, on this latter point, the Director of Public Housing in a large Eastern city announces, in a postscript to the book, a new policy for allocating apartments under his jurisdiction, on the basis of need regardless of race, religion, and color. Thus, the research has already produced a significant social change.

Essentially, the study is a comparative evaluation of two types of public interracial housing projects: "Integrated" (where white and Negro families live as neighbors in the same apartment buildings) and "Segregated" (where Negro and white families live in different buildings or in different parts of the project).

Two projects of each type of occupancy pattern were selected on the basis of their matchability on several factors which experts considered relevant, e.g., racial ratio, age of project, type of neighborhood, ethnic composition of tenants, building structure, etc. About 100 white and 25 Negro housewives in each of the four projects were randomly sampled and intensively interviewed. Also studied by interview was a smaller group of adolescent boys and girls of both races.

The major results are clear cut and significant in the statistical sense. The Integrated Projects in comparison with the Segregated

Projects are characterized by: a greater frequency of friendly contacts between white and Negro tenants, more positive values of housewives with respect to associating with Negroes, a greater amount of neighborly and club activity in general, more positive feelings and respectful attitudes toward Negroes living on the project and toward Negroes at large.

The data are interpreted in terms of three interrelated theoretical processes: the ecology of the Integrated Projects, i.e., the physical and functional proximity of Negro and white families, facilitates more frequent and intimate interracial contacts; the official policy of "integration" is in effect a public sanction for friendly interracial behavior and stimulates the development of social norms which, in turn, direct more friendly racial interaction; the greater discrepancy between the social norms of the Integrated Projects and those of the broader community tends to increase the solidarity in this type of project which produces more favorable attitudes towards the project and its interracial character. The root hypothesis is that friendly, "equal status contact" (a term used frequently throughout the text) of a prolonged and relatively intense nature brings about a diminution of prejudiced feelings and attitudes. Alternative explanations of the data, in terms of the differential selectivity of the tenants in the two types of projects, differences in the projects themselves, and differences in the political, educational, and religious composition of the tenants, are convincingly ruled out.

Despite its strong practical orientation, the book achieves a great deal in the way of integrating some of the complex theoretical interrelationships among ecological, group, and personality factors involved in changing prejudiced attitudes. Some of the theoretical formulations presented, however, require further clarification. (Incidentally, we are told in the preface that a replication of the study is currently being carried out.)

1. While there is little doubt that physical and functional proximity have facilitating effects in regard to interracial contact, the precise conditions under which a contact becomes a neighborly relationship need to be spelled out. Living next door to each other certainly makes it easier for people to become friendly, but is not in itself the "motivating condition" for an intimate relationship.

2. The authors present an excellent discussion of group standards and methods of measuring their effects, but it is not made clear exactly how official policy, implicitly endorsing friendly racial interaction, leads to the development of parallel social norms. The general formulation employed by the authors—that tenants accept the implied interracial standards of Official Housing Management because the latter has the ability to satisfy the tenants' strong need for housing—requires some elaboration at the very least. The theory would have to specify exactly what is implied to tenants by official policy, what particular psychological needs are operative at the time group standards are developed, etc.

3. The notion of "equal status contact," central to the theory of attitude change, is somewhat ambiguous. If it refers to a contact characterized by an absence of preconceptions about Negro inferiority, then it cannot be used as an explanation for reduced prejudice. If it refers to a contact situation in which whites and Negroes have equal rights, authority, and status (as far as the policy and management of the housing project is concerned), it requires some qualification in order to be used as an explanation of reduced prejudice. For, as the authors point out: "an . . . equal status . . . situation which induces competition for scarce goals is likely to promote negative rather than positive sentiments."

Laudable features of the book are its neatness of organization and readable style. Tables are clear and introduced at the right places; footnotes actually amplify as they should, and succinct summaries appear at the end of each chapter. One is conscious of careful editing toward making the thesis of the book readily grasped by non-professionals. A useful appendix includes an illuminating discussion on the nature of prejudice, a few methodological notes (pertaining to the sampling procedure, the training of interviewers), and copies of the interview and rating scales employed in the study.

The book is almost a model illustration of how social scientists can contribute effectively toward the solution of critical problems of our day.

ALBERT PEPITONE.

University of Pennsylvania.

RASHEVSKY, NICHOLAS. *Mathematical biology of social behavior.* Chicago: Univ. of Chicago Press, 1951. Pp. xii+256. $5.00.

The first four chapters of this book summarize previous physiological theorizing and enter, as Rashevsky states, only occasionally into the remaining discussion. The fifth chapter, entitled "Satisfaction and Hedonistic Behavior," is basic to much of what follows. This chapter begins with some speculation as to the physiological nature of satisfaction (pleasure) and proceeds to a discussion of the difficulties involved in its measurement. Satisfaction functions, dependent, in general, upon the pleasure of another individual as well as one's own pleasure, are introduced. Behavior is determined so as to maximize the satisfaction function. The maximum value of one's satisfaction function can be found if the following things are known: (a) all variables which give him pleasure or pain, (b) his degree of egoism, altruism, and sadism (these terms are algebraically defined), toward other people interacting with him, (c) the variables which give these others pleasure and pain, and (d) the method of combination of all of these factors into his own satisfaction function. The remainder of the book is devoted to social hierarchies, imitative behavior, motivational behavior, social learning, and social change. Some unification of treatment is achieved by reliance on the early chapters.

In the entire discussion, Rashevsky relies largely on common-sense notions. The technical vocabulary which has arisen in psychology because of the imprecision of common-sense language is absent. No use is made of the results of laboratory experiments in psychology, nor of the references a psychologist would be apt to cite in connection with some of the topics discussed.

The mathematical treatment is generally simple and fairly straightforward, although its arrangement makes it difficult to winnow out the important conclusions. With the exception of one or two arguments turning on the theory of integral equations, a psychologist with a rudimentary knowledge of the calculus could follow the discussion. This means, of course, that the psychologist who is a fairly expert mathematician will find much of the book somewhat tedious. Considerable attention is given to solutions for equilibrium states and these are situations where the mathematics is usually relatively simple. The mathematical argument is occasionally marred by errors of statement. At the top of page 102, for example, the boundedness of two functions is invoked as justification for proceeding from equation [4] to equation [5]. Reference to page 98 shows that these functions are unbounded. (Consideration of the mean value theorem for integrals shows that the step is justified.)

That most of the topics discussed entail extremely difficult problems relating to the isolation and measurement of the variables involved may be a partial explanation of the paucity of data, experimental or observational, which characterizes this volume. These problems are not, perhaps, problems to be faced within a theory as such, but they loom large when the question of possible verifications is considered. Yet, even within an entirely theoretical discussion, the failure to face such issues sometimes leads to strange implications. For example, Rashevsky states (p. 147): "Finally, communism may be defined as a behavior which corresponds to matrix (5), with sharing determined by the equalization of the satisfaction function of all individuals. This may be considered as a clearcut quantitative definition of the otherwise somewhat vague term 'distribution according to *need*'." (His italics.) It seems obvious that the mathematical precision implied by a definition of this kind is misleading unless there is universal agreement on the form of the satisfaction function. Agreement on satisfaction functions may be as difficult to obtain as agreement on the meaning of the term "need."

Many of the conclusions which Rashevsky reaches are of genuine interest. Many things, known before, are seen somewhat differently in the context of his discussion. Although most psychologists will probably not find this book very rewarding, it may prove highly stimulating and suggestive to social psychologists and to sociologists.

Indiana University. CLETUS J. BURKE.

Miller, D. C., & Form, W. H. *Industrial sociology.* New York: Harper, 1951. Pp. xi+896. $6.00.

Industrial sociology is here "defined as the study of: (1) *work groups and work relations*, (2) *the role the worker plays in work groups*, and (3) *the social organization of work plant society*" (p. 16). Part I begins with an excellent presentation of the classic Hawthorne studies, which are used to demonstrate the need for concern with the social problems of industry. The authors discuss the relations between industrial sociology and such neighboring disciplines as industrial psychology, industrial physiology, and economics.

In Part II the authors squarely face the social problems involved in the relations between labor and management. After a chapter which provides an over-all view of the work society in terms of its material, social, and power organization and of the composition of the labor force, attention is turned to the formal and informal organization of the "House of Management" and the "House of Labor." The concepts of position and role are used in analyses of the activities of the typical "big boss," "plant manager," and first line supervisor. An excellent description of modern unionism is given in terms of the workers organized, the types of unions, the organization of the union local, the union leadership, and the relations between the shop steward and the worker. In describing the social structures of both management and labor, the authors make frequent use of the concept of informal organization, as contrasted with formal organization. Factors associated with societal patterns of power, status, and class are viewed as cutting through both the formal and informal organization of management and labor.

Part III is concerned with the major problems of applied industrial sociology which are divided into two major areas: (1) problems of the job placement of workers, and (2) problems of morale and teamwork. The tools of the industrial sociologist include the interview, the psychological test, the sociodramatic performance test, and the interaction chronograph. It will be of little consolation to the many industrial psychologists who have worked long and hard on the development of psychological tests for industry to learn that "Unfortunately, the variables which have been reduced to precise quantitative measures are not those which have demonstrated a very high correlation with successful job placement" (p. 416). The implication is that sociological techniques have greater promise. While the authors present the results of various studies in an attempt to support this view, this reviewer remains skeptical. A more defensible position would have stressed the value of an interdisciplinary approach to the problems of worker placement, morale, and teamwork.

The social adjustment problems of the worker are analyzed by dividing his life activity into five work periods, and each of these forms a chapter of the fourth major section of the book. The "preparatory work period" includes the minor work experiences of childhood. The "initial work period" extends from the part-time and impermanent job of the schoolboy to the end of formal education. The "trial work period" covers the period from the first full-time job until the individual has found a more or less permanent position. The "stable work period" is the one in which the worker finds a position in which he remains for three or more years. The "final work period" is considered to be retirement. The authors make use of many studies conducted by themselves and their students to illustrate problems of adjustment in each of these periods. This developmental approach also permits a needed integration of many specific studies on occupational choice and occupational interests.

In the final chapters making up Part V, problems associated with the relations between industry and the community and industry and society are considered. These chapters show the interdependence of industry and community in terms of ecological, governmental, and cultural factors, and they stress the facts of social change. The present relations between industry and society are viewed as in a stage of transition characterized by a struggle for power.

Industrial Sociology is not a text for the beginning student, since it assumes that the reader has a basic knowledge of sociology and of social science statistics. But for the advanced student of industrial relations, the numerous footnotes, references, suggested readings, and field work assignments should prove extremely valuable. In addition to its avowed purposes as a textbook, this work represents a significant synthesis of diverse research in the emerging field of industrial sociology.

JOHN K. HEMPHILL.

Personnel Research Board, Ohio State University.

TOMAN, WALTER. *Einführung in die moderne Psychologie.* (Introduction to modern psychology.) Vienna: Humboldt Verlag, 1951. Pp. 266.

This book is unusual in two respects: Written by a European, it displays a thorough knowledge of modern American psychology, and it has a compactness and conciseness which are seldom found in any general text.

An introductory chapter contains an historical sketch of the field. About two-thirds of the space is devoted to contributions since 1900.

The second chapter, which takes up about one-third of the book, describes four "theoretical systems," namely, behaviorism (largely Tolman and Hull), psychoanalysis, "vector psychology" (Lewin), and factor analysis (Thurstone). It is difficult to evaluate how much understanding these condensations might convey to one who is not familiar with the material. However, the care with which the relevant features of each topic are selected and the clarity with which they are discussed are impressive. The author has more than a superficial acquaintance with his sources.

The next chapter deals with the "present status of general psychology." It contains sections on abilities, personality, motivation, affectivity, perception, memory, thinking and the senses. It is no more than a topical outline of these areas, and even as such it is the least satisfactory part of the book. There follow four brief chapters on genetic, clinical, social, and applied psychology. A final chapter, by way of an appendix, contains a description of elementary statistical methods. The name and subject indices are followed by a glossary of English technical terms, which once more characterizes the particular orientation of the author.

On the whole the book accomplishes in a competent manner the aims as expressed in the title. It is an introduction rather than a treatise, and it recognizes that it is the American workers who have contributed the major share to the development of modern psychology.

KARL F. MUENZINGER.

University of Colorado.

PORTEUS, STANLEY D. *The Porteus Maze Test and intelligence.* Palo Alto: Pacific, 1950. Pp. vi+194. $4.00.

The thesis of this book is that the Porteus Maze Test is an indispensable member of a battery of tests for the evaluation of intelligence. In the development of this contention Porteus critically reviews definitions of intelligence, criteria for validation of tests of intelligence, and the methods by which such validation has been attempted. He is especially interested in the study of results of changes consequent to brain surgery. The treatise also includes a manual for the administration, scoring, and interpretation of the Maze Test. Many studies relating to the Maze and its uses are reviewed.

Porteus considers several definitions of intelligence and finds in most of them, although he has to strain a little now and then to do so, the common element of planning. He comes, then, to his own definition of intelligence as "Capacity for making planned responses to an increasing range of relevant stimuli." While such tests as the Binet or the Wechsler are important for measuring, in part, this planning capacity, they are not adequate for the entire task—missing the mark especially with respect to "certain capacities or traits of a practical nature essential to social efficiency." Citations in this connection are made of several accounts of instances in which persons of low IQ's have not exhibited feeblemindedness in the sense of being unable to get along acceptably in their social and economic environments. The Maze is most helpful in the identification of such individuals. The Maze, too, has exceptional value in the estimation of differences of intelligence among racial groups, which must always be evaluated in light of the problems and the environment of the racial group studied. The Qualitative Score, which has been used since 1942, has been shown to separate from the normal those with tendencies towards delinquency. The reviewer finds it very difficult to evaluate, on the basis of the information given in the manual, a performance according to this qualitative score. Wavy lines are supposed to be significant, but there is no good description of "wavy." Examples are given, but these do not show what could be accepted in contrast with what could not.

Evidence for the above claims has been presented before. The new, and for Porteus the final, proof of the value of the Maze comes through brain surgery. The Maze appears to be sensitive to changes in the capacity for planning as commonly observed in lobotomized patients but which are not revealed by such tests as the Binet or the Wechsler. The results of several studies are given. The Maze "has the best claims to being considered a 'frontal brain' test."

Although Porteus is sharply critical of the limitations of widely-used tests of intelligence and of the statements of test users, he states clearly that these tests must always be used along with the Maze.

One detects a rather plaintive note permeating the book. This is understandable in one who for 35 years has been working assiduously with a test of his own making, has seen it used in many studies in many parts of the world, but who still has not seen it accepted as having the primary value he believes it to possess.

Porteus is somewhat repetitious and there are a few instances of careless proofreading. The book is, however, interesting and challenging. Allowing a little for the author's bias, those engaged in clinical practice should read it.

The University of Nebraska. D. A. WORCESTER.

KEYS, ANCEL, BROŽEK, JOSEF, HENSCHEL, AUSTIN, MICKELSEN, OLAF, AND TAYLOR, HENRY L. *The biology of human starvation.* (2 vols.) Minneapolis: Univ. of Minnesota Press, 1950. Pp. xxxii+1385. $24.00.

These two volumes contain a comprehensive report of a broad program of scientific observations on 32 young men during a six-month period of semistarvation, and a roughly comparable period of nutritional rehabilitation. The study had as an over-all goal a scientific evaluation of the problem of food rationing, with the correlated problem of nutritional rehabilitation in war zone famine. Systematic dietary regulation reduced the body weight of subjects by 25 per cent, and the basal metabolism by 40 per cent.

Although data of psychological interest are to be found in many parts of the two volumes, the primary psychological report by Dr. Josef Brožek covers approximately the first half of Volume 2. Measurement techniques applied to the group may be summarized under five categories: (*a*) procedures directed at the assessment of changing status in basic sensory and motor capacities; (*b*) comparison of clinically observed or personally reported effects on mental function with objective performance on two batteries of standardized intellective tests stressing, respectively, speed and power features; (*c*) evaluation of personality changes during semistarvation and rehabilitation by standardized personality scales; (*d*) systematic report of subjective status through interview and complaint inventory techniques; (*e*) general case studies and subject evaluation from an over-all clinical viewpoint.

The progressive circumstances of clinic cases are highly individualized; this fact has emphasized the diagnostic value of personality scales as compared with their potential function as measures of therapeutic progress. On this problem the Minnesota experiment makes a very significant contribution. Scores on the nine scales of the Minnesota Multiphasic Personality Inventory are reported for a control period and for selected points during 24 weeks of semistarvation, 12 weeks of controlled rehabilitation, and a further period required for full return to normal status. As the nutritional stress of the experiment increased, the most significant changes occurred progressively on the scales for Hypochondriasis, Depression, Hysteria, and Psychasthenic and Schizoid tendencies. Promptly on the beginning of the rehabilitation period the direction of this change was altered toward a return to the control level. Demonstration of a controlled test-score lability paralleling a controlled physiological change is rare, and to some degree enables one to regard these tests as dynamic in character, in contrast to views which imply assessment of cumulated psychic residues as a primary function. In one sense these results are perhaps the first example of a physiological validation of a personality scale.

In the intellective area, six tests developed on the basis of Thurstone's factorial studies and organized as a speed test were supplemented by the Thorndike CAVD as a power test. Despite subject complaint of difficulty of concentration, impairment of judgment, and memory difficulties, these test batteries did not reveal any evidence that "intelligence" is adversely affected by the semistarvation routine.

In the sensory area, visual and auditory functions were essentially unaffected by the stress, with some evidence of an increase in auditory acuity. Despite large decreases in the tolerance for physical work and subject reports of coordination difficulties, the primary finding in the motor field was a generally increased threshold for reflex stimulation.

The over-all picture of the results of psychological interest in this exceptionally broad study suggests areas in which new studies might concentrate when and if subjects of sufficient hardihood and experimenters of sufficient patience are found. Now that the important auditory and visual systems are known to be maintained at high level during starvation stress, it would be of decided interest to study the response of olfaction and gustation. The hunger drive is functionally in close relation to taste and smell mechanisms, and investigations of these areas would be of value. In similar manner, a further functionally indicated sensory study would be cutaneous thermal sensitivity.

The solid position of these two volumes in the reference literature of all disciplines related to nutrition is assured, and in view of the great importance of the drives and motivations connected with nutritional mechanisms, the psychological contribution is one of the departments of major interest. The report will appeal especially to physiological psychologists and to those concerned with motivation problems at either the human or the animal level.

L. P. HERRINGTON.

Yale University and
The John B. Pierce Foundation.

LUCAS, D. B., AND BRITT, S. H. *Advertising psychology and research.* New York: McGraw-Hill, 1950. Pp. xi+765. $6.50.

A book with the scholarship and scope of this one has long been overdue in the field of psychology in advertising. No book of recent vintage has encompassed such a wide variety of topics with such adequacy. In the main, this reviewer has had to refer students to original references. With the publication of *Advertising Psychology and Research*, the necessity of this practice will be substantially reduced.

The few bad aspects are overwhelmed by the large number of excellent features. Above all, the authors have been candid. They have not glossed over weaknesses in methods or in results. Although the results of laboratory studies are presented, the authors repeatedly warn the reader of the dangers of extrapolation from laboratory situations, which are overcontrolled to a point of artificiality, to the vastly more complex day-to-day advertising world. To a greater degree than has been done before, they draw upon a wealth of material obtained by the field survey and its more adequate sampling. To be sure, there is less "single variable" control, but there is the compensation of an approach to reality. Much material is reported from non-journal sources—from commercial studies, privately published pamphlets, and from unpublished research available to the authors.

Another laudable feature is the authors' restraint in unwarranted generalization either by word or implication. There are many fine reproductions of advertisements, but these are not presented as ideals or "successful ads because." Rather, they are presented as illustrations of different methods used in attempts to improve advertising effectiveness. The authors have resisted the temptation to quote anecdotal material in favor of more "solid stuff."

In general, the exposition is lucid and interesting. The chapters on methods are somewhat difficult, but this is not the authors' fault. The subject matter is difficult. Tables are profuse and, in the main, lead to easy interpretation and emphasis. Examples of commercial radio research reports are particularly well described. The format is attractive and contributes to the over-all favorable impression of the book.

It is with the organization of content that this reviewer would find fault. The first four sections discuss the conventional topics of attention, memory, motivation, and mechanical factors involved in advertising effectiveness. The last two sections are devoted to techniques of measurement. Logically and practically, a reversal of this arrangement would be in order. One should be exposed to methods of obtaining data before sampling the data themselves. Evaluation of results should partly depend upon the techniques used to gather the data, and this is especially important when different methods are brought to bear on a problem. The advertiser is often skeptical of research yielding apparently conflicting reports, because he does not realize that different methods measure different things. Moreover, early presentation of techniques would eliminate references to methods that are not clarified until somewhat later in the book.

The authors occasionally stray from a strict use of terminology, such as "heard a program" rather than "had the radio turned on." There is an understandably strong bias in favor of the controlled recognition technique. The authors might have included more critical appraisal and suggestions for further research at some points.

In view of the excellent job done by the authors in a rather difficult subject matter area, this reviewer feels certain that *Advertising Psychology and Research* will have wide acceptance by the advertising practitioner and by teachers.

LESTER GUEST.

The Pennsylvania State College.

BLOOM, B. S., AND BRODER, L. J. *Problem-solving processes of college students; An exploratory investigation.* Supplementary Educational Monographs, No. 73. Chicago: Univ. of Chicago Press, 1950. Pp. 109.

The authors of this monograph sought to find out how college students deal with examination problems and to develop a useful remedial method. They emphasize that simple scores of success or failure often do not adequately measure actual performance.

The study utilized problem-items typical of comprehensive examinations at the University of Chicago. Protocols were obtained from small groups of students, who were asked to "think aloud." A brief training period in this technique preceded exposure to the crucial problems.

Topics covered include analysis of individual differences, comparison of students academically successful and unsuccessful (six students in each group), the nature of problems, and remediation of problem-solving. Special attention is devoted to a group who failed examinations but whose failure did not seem to represent their real abilities. After analysis of particular difficulties, mainly by comparing the group with "models" (successful problem-solvers), the students practiced typical problems in order to acquire the model students' methods. Training resulted in significant improvement in examination grades. The most evident change was an increase in confidence.

The review of the literature on problem-solving is very inadequate—only passing reference is made to Wertheimer and Maier, and no mention at all of Claparède, Katona, Duncker, Durkin, Luchins, or Woodworth. Culling of available material might have enabled the authors to avoid some naive statements and have given them valuable leads for their own work. As it is, their remedial program depends almost entirely upon a preliminary study with a very few cases.

Largely subjective and testimonial techniques are used to assess the remedial program. Despite improvement in grades, little evidence is given of genuine changes in reasoning. Better grades, in fact, might be a function of improved motivation.

The authors deserve credit for having elucidated a neglected point, that processes of problem-solving are as significant as the end-product. They well emphasize that training can be given, not alone in how to study, but also in how to think. Their analysis of problems is good, and they contribute some interesting hypotheses for further research, especially those pertaining to "foreground-background" factors in problem-solving.

University of Hawaii. W. EDGAR VINACKE.

HARTLEY, E. L., BIRCH, H. G., & HARTLEY, RUTH E. (Eds.) *Outside readings in psychology*. New York: Crowell, 1950. Pp. xiii+875. $2.75.

There can be no doubt of the fact that this collection of readings is a bargain. But a bargain for whom? The editors have made available at minimum cost an imposing number of highly important contributions to psychological fact and theory. But availability at low cost is scarcely a sufficient objective in either science or education. For whom and to supplement what are these "Outside Readings" intended? The editors declare they are aimed "to complement standard textbooks in introductory psychology," and this objective is confirmed by presenting on page 865 a table to indicate parallels between these readings and the appropriate chapters in Boring, Langfeld, and Weld, Dashiell, Munn, Ruch, and Woodworth and Marquis. However, if Lashley's article on "Coalescence of Neurology and Psychology," from the 1941 Proceedings of the American Philosophical Society is appropriate collateral reading for the students in the introductory course in psychology at the College of the City of New York, this reviewer is profoundly impressed by their selection of students. The editors do add, to be sure, that "it is anticipated that the wealth of material included will lead to wide acceptance of this volume in many intermediate and advanced courses as well." The volume deserves such acceptance.

The collection comprises ninety-eight readings, grouped under seventeen such customary introductory textbook headings as Development, Perception, Learning, Emotion, Social Behavior, and the like. In any such mass of material the selection must inevitably be uneven in quality. In the opinion of this reviewer the selections in the sections on Development and on Intelligence are the most satisfactory, although Spelt's article on "The Conditioning of the Human Fetus *in Utero*" would seem to be too technical for the beginning student, and the Kornhauser article "Are Intelligence Tests Worth While?" is suggestive of "Methinks the lady doth protest too much." The four readings which constitute the Introduction impart no information of any consequence. Two of the four in the section on Attention and Set discuss a theoretical issue of no interest to the beginning student. The separation of Perception and Sensation in two distinct sections is a questionable procedure. Under the heading of Motivation, Cannon's 1929 discussion of thirst is dated. And it is difficult to conceive of the function of an "outside reading" in statistics which comprises a seven-page summary of procedures and formulae.

For the student in an Advanced General Psychology course most of these readings are appropriate and valuable, as they may also be for the graduate student preparing for comprehensive examinations, or for the undergraduate tutorial student. For the beginning student, however, they require clarification and elaboration which the instructor in most introductory courses is not likely to have the time to devote to "outside readings." For example, the article by Hebb on "Man's Frontal Lobes" is an excellent one, but is essentially clinical and mentions a great many clinical tests which must be explained to the student if he is to understand the reading, with the result that an inordinate amount of class time may be devoted to a discussion of techniques alone. The fact that most of these readings originally appeared in the professional psychology journals would seem either to preclude many of them being appropriate, in their original form, for the beginning student, or else raise an interesting question concerning the level of content of our professional journals.

CARLETON F. SCOFIELD.

University of Buffalo.

PIAGET, JEAN. *The psychology of intelligence.* New York: Harcourt, Brace, 1950. Pp. viii+182. $3.50.

In the short time since its first publication, this popular book has already passed through a second printing. Fortunately, it is now available in an English edition. Unlike most contemporary books concerned with intelligence the present one contains only one small reference to intelligence tests as such. No mention is made of the recent activities of such factor theorists as Thurstone, or of Terman, Wechsler and others in America who have contributed much to the mental testing movement this past quarter century. To some psychologists this may be somewhat surprising and perhaps disappointing, but to those who know of Piaget's earlier writings it will appear quite consistent.

Part One consists of two chapters, the first of which is devoted to a discussion of the nature of intelligence, its relationship to adaptive processes in general, and a historical review of the possible interpretations of intelligence according to the various psychological and philosophical systems. This is followed by a carefully organized analysis and critique of Thought Psychology, in particular the relation of logic to intelligence. Starting with logic as the axiomatics of reason and the psychology of intelligence as the corresponding experimental science, Piaget develops his concept of operations and their "groupings," a construct which is essential to his entire theory of intelligence. Briefly, operations are actions which have been internalized and organized in complex structures. A grouping is a special equilibrium or configuration of operations, one which meets at least four conditions: (1) any two elements of a grouping may be combined, resulting in a new element of the same grouping; (2) every change is reversible; (3) the combination of operations is "associative" in the logical sense of the term; and (4) an operation combined with its converse is annulled. This last condition refers to the conservation of the whole. For example, in young children Piaget finds that a return to the starting-point is not accompanied by a conservation of the latter. The primary task for a psychology of intelligence is to describe this "grouping" in relation to the various genetic levels which lead up to it, and in contrast to forms of equilibrium characteristic of other functions such as perception.

The second part of the book deals with the relationship of intelligence to perception and habit. None of the four major conditions of "grouping" which characterize intelligence is realized at the level of perceptual structures. Perception proceeds by immediate but partial contact with its object, distorting it by the very act of centering or focusing upon it, although these distortions are reduced by partial decentralizations. Intelligence, on the other hand, encompasses in a single whole a much larger number of facts reached by variable and flexible paths, and attains objectivity by a much more thorough decentralization. Intelligence is thus the ideal "grouping" toward which perceptual structures tend. In a similar manner, Piaget differentiates between habit and intelligence, pointing out how sensori-motor (preverbal) intelligence develops from assimilatory activity. His penetrating analysis of the construction of the object and of spatial relations as well as size and shape constancies is particularly thought-provoking and deserves careful study.

Part Three is chiefly concerned with the growth of thought processes from the preconceptual stages to the highest level of formal operations. In a concluding chapter Piaget attempts to relate his theory of intelligence to a universal succession of fundamental structures inherent in life itself, rhythms followed by regulations which in turn ultimately become "groupings" as manifest in the intelligence of man.

Although numerous observations are cited to support various arguments, many psychologists will no doubt quickly grow impatient with the lack of sufficiently documented experimental evidence. Others will find this presentation highly abstract and difficult to grasp on first reading. But those who take the time to gain a thorough mastery of this little book will be richly rewarded by gaining new insight into a theory which took Piaget a lifetime to formulate.

WAYNE H. HOLTZMAN.

The University of Texas.

DORCUS, ROY M., & JONES, MARGARET HUBBARD. *Handbook of employee selection.* New York: McGraw-Hill, 1950. Pp. xv+349. $4.50.

This should be a useful reference book for industrial psychologists, personnel executives and others concerned with personnel selection and placement on the basis of psychological tests. It consists of noncritical abstracts (mostly from original sources) of references dealing with the selection of employees by means of psychological tests. Excluded are studies utilizing students in trade, commercial, and professional schools; military and aviation studies; and studies of accident proneness.

The authors examined over 300 different periodicals and many books and monographs. Out of 2,100 references examined, 427 were selected for inclusion in this volume. A few of the references are from foreign publications. The usefulness of these latter is somewhat open to question. It is difficult to judge the applicability of a specific procedure or test battery from one industrial situation to another; this difficulty is markedly increased when we attempt to make such a judgment from one country to another. However, it is of interest to the personnel research worker to see what has been done outside of the United States in personnel testing.

The particular selection of references would, of course, not be wholly satisfying to every reader. Studies conducted by such competent workers in the field of industrial psychology as Ghiselli, Harrell, Tiffin and Viteles are well represented. One feels a bit uncomfortable, however, in the presence of the apparently glittering results obtained by Chapple with his Interaction Chronograph, results which other workers have failed to duplicate following Chapple's procedures. In instances of this sort one wishes that the authors had not adhered so uniformly to their objective of noncritical abstracting.

The general format of the book is good. Abstracts are presented chronologically by publication date and numbered. Each abstract includes, in addition to bibliographical data: number of subjects, types of jobs under investigation, tests used, criterion, validity, and reliability (when available). Three good indices should aid the user of this book in finding his way around. There is a job index at the front of the book; test and author indices are at the end.

A rough count yields the following distribution of references by publication date: 1900–9, .2%; 1910–19, 3%; 1920–29, 28%; 1930–39, 26%; 1940–48, 43%.

This book meets a need of workers in personnel psychology, particularly those in business and industrial concerns who do not have constant and immediate access to a well-stocked library. Also, as the authors point out in the Preface, "literature in this field is so widely scattered that even in the libraries of large universities it is impossible to find many of the references." A real service has been done in bringing these source materials together.

CAROL S. BELLOWS.
ROGER M. BELLOWS.
Wayne University, Personnel Research Center.

BRAYFIELD, ARTHUR H. *Readings in modern methods of counseling.* New York: Appleton-Century-Crofts, 1950. Pp. xii+526. $5.00.

Brayfield's source book of forty-seven articles represents a major attempt to present in readily available form a sample of some of the more important research and thinking which has appeared during the past ten years in the field of counseling with normal individuals, particularly with college students. The articles are roughly divided into six categories covering counseling trends, the clinical method, diagnosis, treatment, interviewing, and evaluation.

As one reads the collected articles, the more or less confused state of counseling, the differences in opinions, the absence of very many facts, the difficulties of carrying out significant research, and the need for more carefully controlled research all become disturbingly apparent. Whether the effect on the general student or on the counselor in training will be one of bewilderment, frustration and discouragement, or whether the effect will be to urge on these students and other workers to contribute their bits to this important field, only experience in the use of the text will tell. It seems likely that the really effective instructor who has a sufficiently sound theoretical background and sufficiently broad practical clinical experience will find this book very useful at least as a supplementary text. He would be able to help the students organize, systematize and evaluate the various articles in such a way that a great deal of the possible confusion in the beginning students' thinking would be minimized. A somewhat lesser burden would have been placed on the instructor if Brayfield had allowed himself more space at the beginning of each of the six major sections of the book for providing a general criticism and integration of the different articles.

In contrast to the more general texts on counseling, this book does introduce the student to a wide variety of significant and interesting research articles. It should consequently serve to stimulate research interest provided the book is used properly. It is doubtful, however, that it can or should serve as a substitute for the more concentrated reading of current research articles which a serious graduate student would need to do in a well-supplied library. It also seems unlikely that Brayfield's book will serve as an effective substitute for the more general texts on counseling and psychotherapeutic procedures. Despite the editor's emphasis on what he terms "process" and "method" it is questionable how much a student would obtain which would be of real use in conducting a counseling interview. The compilation does not present any consistent and integrated picture of the counseling process.

The general selection of articles for inclusion seems to have been quite carefully and thoughtfully done. The articles include a wide range of subject matter and a variety of viewpoints. In general, the more significant articles have been chosen. There is a preponderance of articles stemming from Rogers' viewpoint and that of the Minnesota group. Brayfield argues that this simply reflects the greater number of articles published by representatives of these groups. Nevertheless, there does seem to be unwarranted neglect of important theoretical articles relating counseling and therapy to the learning process, of material describing systems of personality organization and therapy, and of the contributions of the analytic and neo-analytic schools. This is somewhat at variance with Brayfield's recognition of the growing tendency to associate counseling and therapy and of the importance of the counseling relationship.

Like any collected sample of articles, this book is extremely difficult to evaluate generally. It is possible that for some instructors in some courses it could serve as a primary text. To the reviewer it represents quite a good and readily available source of supplementary reading for beginning students in counseling or for introducing the non-specialist—the mature faculty counselor, for example—to the complexities involved in the counseling process.

WILLIAM M. GILBERT.

University of Illinois.

ABT, LAWRENCE E., & BELLAK, LEOPOLD. *Projective psychology.* New York: Knopf, 1950. Pp. xii+485. $6.00.

Projective psychology, like many other of our applied sciences, seems to be following a developmental path that may be categorized into three stages. In the first of these a gadget is discovered which seems to work. The practitioner doesn't know why it does, and except for minor variations is unable to improve it. In the second stage, scientific investigation of the gadget and the phenomena on which it operates yields testable hypotheses based on an integrated theoretical structure. In the third stage, the knowledge gained from the experimental testing and constructions in stage two permit evaluation of existing techniques, identification of meaningful operations, and the construction of more effective instruments.

The publication of this volume suggests that the projective study of personality is well into its second stage of development. Siple's bibliography, which was originally to have appeared in this book, contains some fifteen hundred titles. Of these, only twenty-seven deal with theories as to how projective techniques operate—and why they seem to work. Previous volumes on these techniques have, for the most part, simply described individual tests.

Part One consists of two theoretical papers by Bellak and Abt. Their contribution to this rapidly developing field of psychology is an attempt to present a testable theory of projection. Since it is testable, the tenability of their hypotheses, which derive from an integration of certain concepts in psychoanalysis and gestalt psychology, can be determined by direct experimental work. Bellak reports a few initial studies, and it is to be hoped that others will undertake further testing of these hypotheses.

Part Two consists of ten papers describing eight different tests. Three papers on the Rorschach (content analysis, inspection technique and group techniques) are followed by descriptions of the TAT, the Mosaic Test, Finger Painting and Finger Drawing, the Szondi, and the Bender Visual-Motor-Gestalt Test. The authors are well known clinicians with considerable experience in the use of the instruments they discuss. The quality of these papers is, however, highly uneven. Except for Bellak's paper on the TAT, none of the authors relates his discussion to the hypotheses in Part One. Only three present data on sizable groups at all useful in evaluation of the instruments they describe; four present clinical data only, or studies of small samples such as in Deri's study of Szondi Test changes in nineteen depressives before and after electric shock therapy, with two control groups of ten Ss each. The other papers are either theoretical reviews of the literature or descriptions of tests and how they are scored. The erratic quality of this second section is in part compensated for by the fact that there are some tests described in detail here which have not previously been included in such collections.

Part Three consists of two papers—one on the use of projective tests in business and industry, and the other on their use in action research. These are both reviews of selected literature with some examples of applications. While not as complete as the reviews by Campbell and by Deri, Dinnerstein *et al.*, which have appeared in this Journal, they are useful supplements to them.

This is not an instructional manual like Rapaport, Gill, and Schafer's *Diagnostic Psychological Testing;* it is considerably more detailed in its description of fewer tests than is Rosenzweig and Kogan's *Psychodiagnosis*, and it does not attempt the bibliographic coverage of Bell's *Projective Techniques*. It does present a coherent theoretical structure, descriptions of eight specific tests, and an introduction to their use in a variety of situations.

University of Illinois. IRVING LAZAR.

MASLOW, A. H., & MITTELMANN, BELA. *Principles of abnormal psychology.* (Rev. Ed.) New York: Harper, 1951. Pp. xv+665. $5.00.

The revised edition of this text is somewhat larger than the original, but it retains most of the factual content and writing of the first edition. Like the earlier edition, its general orientation is psychoanalytic, but its presentation is non-doctrinaire, and it successfully synthesizes findings from clinical, experimental, sociological, psychobiological, and anthropological research. The best features of the first edition are also found in the revision: emphasis on motivational roles of symptoms; clearly defined dynamic concepts; lucid accounts of the role of culture in determining personality structure; clarification of the correlations between somatic and psychological aspects in psychopathology; excellent descriptions of the major abnormal reaction types, enlivened by well-documented case histories.

This edition contains several new chapters and sections. For example, there is a competent discussion of somatic treatment and a chapter on "Reactions to Conditions of War" which is an excellent summary of the work of Grinker and Spiegel. The new chapters on psychopathic personality and character disorders are worthwhile additions to the book, the latter including a particularly good discussion of the relationship between these disturbances and the syndromes of more severe behavior deviations. The section on psychotherapy has been considerably expanded and made more comprehensive. A noteworthy new section on "Syndromes in Infancy, Childhood, and Adolescence" reviews the significant work of Bender and many of the important contributions which have been reported in the several volumes of *The Psychoanalytic Study of the Child.*

Some of the new features, such as the brief section on extrasensory perception and the chapter on preventive techniques in school and industry, seem superfluous or inappropriate. Uncritical discussions of the projective techniques, both in the text and in the appendix, may lead the reader to believe that these techniques have greater validity than systematic studies have actually demonstrated.

A major change in point of view on the basic issue of genetic vs. environmental determination of abnormality has occurred. In the original edition, discussions of etiology revealed a marked bias in favor of environmental explanations. In the revised edition the authors state that Kallman's recently published data "*convincingly establish* the presence of a genetic factor in . . . schizophrenia" (p. 119; italics ours). They also conclude that genetic factors play an important role in manic-depressive psychosis (p. 120). This reviewer doubts the wisdom or necessity of such a radical shift in point of view, particularly in the light of the critiques of Kallman's work and the increasing number of studies —some of them summarized in the text—which forcefully demonstrate the relationships between very early parent-child relationships, emotional development and later personality disorders. The authors do not consistently maintain a genetic or constitutional point of view, stating in a later section of the book that "the affective and schizophrenic [reaction types] are assumed to be chiefly psychogenic" (p. 501).

It is the reviewer's opinion that the major shortcoming of the book is its omission of some outstanding recent developments in theory and experimentation in the field of psychopathology. For example, although the authors give excellent accounts of the *meanings* of symptoms to patients, there is no discussion of the *learning* of these symptoms. Cameron's profound conceptualizations of social learning and role-taking in relation to self-evaluations, defense mechanisms, and symptom formation are completely neglected.

The rigorous formulations of conflict theory and the excellent experiments derived from these formulations by J. S. Brown and N. E. Miller are not cited. There is no mention of the significant contributions of Mowrer and of Dollard and Miller which have furthered our understanding of abnormal psychology by applying learning theory and experimentation to psychopathological and psychotherapeutic phenomena. Important objective and experimental studies of schizophrenic language and thought and of the effectiveness of psychotherapy are omitted. The brilliant work of the Chicago Institute for Psychoanalysis is distinctly underplayed, while nondirective therapy is treated in a naive and perfunctory manner.

Probably no textbook could review the vast research and theory in a field as complex as abnormal psychology. Despite its shortcomings, this book ranks very high among the available texts in the field. Together with supplemental readings and lectures to compensate for its deficiencies, the book can serve well as the core of a thorough and stimulating course in abnormal psychology.

Ohio State University. PAUL H. MUSSEN.

HORNEY, KAREN. *Neurosis and human growth.* New York: Norton, 1950. Pp. 391. $3.75.

After many years of slashing at Freudian hypotheses, Horney has at last come up with some positive contributions of her own. Unfortunately, these contributions appear to be not a whit less unscientific or mystical than the Freudian dogmas she has been vigorously combatting.

Horney's latest position is that every individual has several distinct selves, including: (1) A *real self*, consisting of "that central inner force common to all human beings and yet unique in each, which is the deep source of growth"; (2) an *actual self*, consisting of what the person truly is, with all his good and bad, weak and strong traits; (3) an *idealized* *self*, or *pride system*, consisting of the individual's picture of himself as a great and glorious person who can do no wrong and has no weaknesses; and (4) a *hated self*, consisting of the individual's self-contempt and self-alienation.

The insecure individual, says Horney, refuses to face his actual self; builds up a pride system and an idealized self; unconsciously senses the difference between his actual self and idealized self and comes to hate himself for this difference; and, finally, his self-hatred develops into self-conflict against "the emerging constructive forces of the real self"—thus leading to the *central inner conflict*, which is "between healthy and neurotic, constructive and destructive forces."

Some of the serious criticisms of her views may be stated as follows:

1. Horney's *real self*, or "central inner force . . . which is the deep source of growth," is an entirely hypothetical construct which seems to be the optimistic counterpart of Freud's pessimistic death instinct. Like hypotheses concerning the human soul, spirit, elan vital or fate, this hypothesis can never be factually validated or disproven; hence it is scientifically meaningless and redundant.

2. Her *real self* is tautologically defined in terms of man's innate tendency to "grow" or "realize" himself. It means, in plain English, that (*a*) anyone who is born a man must live as a man—and not as an ape, a kangaroo, or a Martian; and (*b*) that an individual's *real self* strives for self-realization—which is certainly the worst kind of psychological tautology.

3. Horney sees an individual "growing" or "realizing himself" only "if he assumes responsibility for himself . . . is truthful to himself . . . is active and productive . . . (if) he relates himself to others in the spirit of mutuality." These are obviously value judgments approved in our own society but not necessarily in other societies. Horney's attempt to establish her *real self* on a biological, innate basis is most surprising in the light of her persistent criticisms of Freudianism on cultural grounds.

4. Horney's theory of neurosis can easily be rewritten in terms of the individual's *pride system*, or *idealized self* being in serious (conscious and unconscious) conflict with his *actual self*. The actual self—meaning the totality of the individual's behavior—indubitably exists, whereas the *real self* is an arbitrary, hypothetical concept.

5. To say that a person's inner psychological drive toward self-realization (*real self*), has been overcome by his drive toward psychological destruction (*hated self*), and that therefore he has become neurotic, is like saying that his unhealthy self has overcome his healthy self. One new hypothesis and two new terms—but no new knowledge—are thus added to our language.

6. Horney implies that man's self-constructive forces are innate, while his self-destructive forces are learned and artificial. Many of the Freudians imply just the opposite. As yet, there is no factual evidence to support either of these extreme views. It is much more reasonable to assume, in our present state of knowledge, that man's constructive *and* destructive drives are determined by innate *and* learned conditions.

7. By breaking down whole organisms into several different kinds of *selves*, Horney resorts to atomistic, unrealistic, and antiquated thinking which does great violence to modern holistic concepts of human personality.

In spite of these serious limitations, *Neurosis and Human Growth* seems to reflect a sound and trenchant understanding of conscious and unconscious dynamics. It is unfortunate that in her attempt to set up highly original, anti-Freudian concepts the author has fallen into such a welter of unscientific, hypothetical, undefinable, moralistic, and metaphysical verbiage.

ALBERT ELLIS.

New Jersey Dept. of Institutions and Agencies.

REDL, F., & WATTENBERG, W. W. *Mental hygiene in teaching.* New York: Harcourt, Brace, 1951. Pp. xiii+454. $5.00.

This is an attempt to convey to teachers in non-technical language the broad aims of a mental hygiene program in the school, and to suggest specific means for achieving them while carrying on the necessary activities of a teacher. The authors succeed in this very well. They present some fundamental principles of dynamic behavior and of personality development and distortion. They are well acquainted with classroom activity and discuss specific means of handling these situations so as to provide a wholesome psychological atmosphere.

The book is not academic. Many concepts and ideas are introduced but always in relation to their meaning for understanding the child and providing a wholesome social climate for his growth. Most of the well-known factors which influence personality development are presented with illustrations and with occasional documentation. There is no attempt to survey the literature. The readings provided are relevant to the background of the teacher. There is a full chapter on group life in the classroom, with specific illustrations of the implication of concepts such as sub-groups, role, group atmosphere and dynamics.

A chapter is devoted to teachers' problems, with specific suggestions. One appendix presents brief discussions of periodicals, pamphlets, audio-visual aids, national organizations and local resources that should be of help to the teacher. Another appendix consists of definitions of special terms which are given in order to prepare the reader for scientific terminology encountered elsewhere.

Those who expect to find a book suited to the training of psychology majors will be disappointed. This text meets the realities existing in the needs and background of the classroom teacher who is taking college courses.

University of Missouri. FRED MCKINNEY.

WEPMAN, JOSEPH M. *Recovery from aphasia.* New York: Ronald Press, 1951. Pp. xii+276. $4.50.

Dr. Wepman's primary purpose has been to provide an orienting manual of aphasia therapy for speech therapists, neuropsychiatrists, psychologists and others whose professional duties may bring them into contact with the brain-injured. He has succeeded in producing an interesting and readable text which should admirably fulfill its mission. The first part of the book is devoted to a discussion of classification, symptomatology and the non-language behavior characteristics of the brain-injured patient.

A short review of etiological factors leads Wepman to adopt a "nonlocalizationist" point of view dictated, as he indicates, more by therapeutic aspirations than by a rigorous scrutiny of available factual evidence from surgery and autopsy table. As Wepman puts it, "While not overlooking the contribution of the 'localizationists,' the writer believes that a more hopeful prognosis can be made for aphasic adults with the acceptance of a nonlocalizationist viewpoint in accordance with which recovery follows reintegration of the remaining cortical tissue into a functioning whole" (p. 18). Here, of course, we are confronted by the fact that a number of quite specifically localized cortical lesions are followed by aphasia, and that retraining by skilled therapists restores some or all of the lost function. Equipotentiality and localization of function in specific spatial loci are not incompatible.

Wepman could have strengthened his theoretical discussion of etiology by grappling with problems of hemispheric dominance in the etiological chapter rather than in his later discussion of non-language factors. He notes (but does not point out its significance for localization) that by far the majority of aphasics have left hemisphere lesions, lesions of the right hemisphere producing rather more dysarthria with little clear-cut evidence of genuine "propositionalizing" aphasia. As Wepman says, if further observation should bear out this left-brain dominance, "the implications appear to be far-reaching and important"—for our knowledge of functional localization.

The major portion of the book is concerned with the details of aphasic therapy. The working principles Wepman sets down seem to make good psychological common sense. The patient is to be considered as a personality whose motivational and emotional life have been altered by the circumstances of his injury and disability, whose therapy should take place in a climate of understanding of these personality factors and should be initiated as soon as possible after the injury. Of interest to the clinician and physiological psychologist alike in this section should be the description of the self-concept of the aphasic patient. Impressive is the extent to which the deprivation of language function and the attendant stress of injury reduce the patient's behavioral level to something resembling early infancy. Further investigation of the self-concept of those brain-injured patients who are not aphasic is necessary, of course, before these data can be evaluated adequately.

The discussion of techniques achieves a happy compromise between specific procedural details and their underlying reasons. Although

Wepman appears to use a consistent Gestalt approach towards learning theory, his actual recommendations for practice contain many details of stimulus repetition and reward for success that should be suggestive to other varieties of learning theorist. This section of the book contains a host of suggestions for collaborative research by experimentalist and speech therapist.

Recovery from Aphasia, more than any book heretofore available, should meet the needs of training programs in clinical psychology and psychiatry which require a non-specialized but authoritative account of therapy in aphasia. The value of this book to theoretical and physiological psychologists is rather less in the formal presentation of etiological fact and theory than in the many problems for research in aphasia which the author's wide practical experience suggests.

University of Illinois. L. I. O'KELLY.

BOOKS AND MATERIALS RECEIVED

Arnold, Wilhelm. *Das Raumerlebnis in Naturwissenschaft und Erkenntnistheorie.* Nürnberg, Germany: Sebaldus, 1949. Pp. 176.

Asch, Morton Jay. *Nondirective teaching in psychology: An experimental study. Psychological Monographs,* No. 321 (Vol. 65, No. 4). Washington, D.C.: American Psychological Assn., 1951. Pp. iii+24. $1.00.

Barker, Roger G., & Wright, Herbert F. *One boy's day; A specimen record of behavior.* New York: Harper, 1951. Pp. x+435. $3.50.

Bierer, Joshua. *The Day Hospital: An experiment in social psychiatry and syntho-analytic psychotherapy.* London: H. K. Lewis, 1951. Pp. 56. 6*s.*

Bohm, Ewald. *Lehrbuch Der Rorschach-Psychodiagnostik.* Bern: Verlag Hans Huber, 1951. Pp. xxiii+406. Ganzleinen Fr. 36.-

Bugelski, B. R. *A first course in experimental psychology.* New York: Holt, 1951. Pp. xxiii+421. $3.50.

Coleman, James C. *Library of test items to accompany "Abnormal Psychology and Modern Life."* Chicago: Scott, Foresman, n.d. Pp. 85.

Crow, Lester D., & Crow, Alice. *Mental hygiene.* (2nd ed.) New York: McGraw-Hill, 1951. Pp. xviii+433. $4.50.

Culbertson, James T. *Consciousness and behavior: A neural analysis of behavior and of consciousness.* Dubuque, Ia.: W. C. Brown, 1950. Pp. xii+210. $4.25.

Davey, Harold W. *Contemporary collective bargaining.* New York: Prentice-Hall 1951. Pp. xi+532. $8.00.

Davidoff, Eugene, & Noetzel, Elinor S. *The child guidance approach to juvenile delinquency.* New York: Child Care Publications, 1951. Pp. vi+173. $4.50.

Edwards, Allen L. *Psychological scaling by means of successive intervals.* Chicago: Univ. of Chicago, Psychometric Laboratory, 1951. (No. 69.) Pp. 13.

English, Horace B. *Child psychology.* New York: Holt, 1951. Pp. xiv+561. $4.50.

Feingold, S. Norman. *Scholarships, fellowships and loans.* Vol. II. Boston: Bellman, 1951. Pp. 312. $5.00.

Gemelli, Agostino. *La strutturazione psicologica del linguaggio studiata mediante l'analisi elettroacustica.* Scripta Varia: Pontificiae Acad. Sci., 1950, No. 8.

Jeffrey, Thomas E., & Wegener, Jonathan. *Response timer.* Chicago: Univ. of Chicago, Psychometric Laboratory, 1951. (No. 70.)

Jones, Lyle V. *Primary mental abilities in the Stanford-Binet, age 13.* Chicago: Univ. of Chicago, Psychometric Laboratory, 1951. (No. 71.) Pp. 19.

GARRETT, HENRY E. *Great experiments in psychology.* (3rd ed.) New York: Appleton-Century-Crofts, 1951. Pp. xvii+358. $3.50.

GUETZKOW, HAROLD. (Ed.) *Groups, leadership and men: Research in human relations.* Pittsburgh: Carnegie Press, 1951. (New Brunswick, N. J.: Rutgers University Press, authorized distributing agent.) Pp. ix+293. $5.00.

KATONA, GEORGE. *Psychological analysis of economic behavior.* New York: McGraw-Hill, 1951. Pp. ix+347. $5.00.

KITSON, H. D. *How to use your mind.* (4th ed.) Philadelpha: Lippincott, 1951. Pp. 196. $2.50.

KUDER, G. FREDERIC. *Examiner manual for the Kuder Preference Record—Vocational—Form C.* Chicago: Science Research Associates, 1951. Pp. 32. $.75.

KUDER, G. FREDERIC, & CRAWFORD, LURA E. *Kuder Book List.* Chicago: Science Research Associates, 1951. Pp. 8. (Package of 25, with Instructor's Guide—$2.35.)

LECKY, PRESCOTT. *Self-consistency: A theory of personality.* New York: Island Press, 1951. Pp. v+275. $3.50.

MACHOVER, KAREN (Chm.) *Case reports in clinical psychology.* Brooklyn: Kings County Hospital, 1950. (Vol. I, No. 2, pp. 71, $1.00; Vol. I, Nos. 3 & 4, pp. 52, $1.00; Vol. II, No. 1, pp. 78, $1.25.)

MCCLELLAND, DAVID C. *Personality.* New York: William Sloane, 1951. Pp. xvi+654. $5.50.

MASLOW, PAUL. *The individual through the Rorschach.* Vol. I of The Life Science. Brooklyn, New York: Author, 1951. Pp. 134. $3.50. (Mimeo.)

MILLER, JAMES GRIER. (ED). *Experiments in social process: A symposium on social psychology.* New York: McGraw-Hill, 1950. Pp. ix+205. $3.00.

MORENO, J. L. *Sociometry, experimental method and the science of society: An approach to a new political orientation.* Beacon, New York: Beacon House, 1951. Pp. xiv+220. $6.00.

MURSELL, JAMES L. *Streamline your mind.* Philadelphia: Lippincott, 1936. Pp. 253. $2.95.

MURSELL, JAMES L. *Using your mind effectively.* New York: McGraw-Hill, 1951. Pp. ix+264. $3.00.

PASCAL, GERALD R., & SUTTELL, BARBARA J. *The Bender-Gestalt Test: Quantification and validity for adults.* New York: Grune & Stratton, 1951. Pp. xiii+274. $6.50.

PIMLOTT, J. A. R. *Public relations and American democracy.* Princeton: Princeton Univ. Press, 1951. Pp. xi+265. $4.00.

RAPAPORT, DAVID. *Organization and pathology of thought: Selected sources.* New York: Columbia Univ. Press, 1951. Pp. xviii+786. $10.00.

SCHNEIDERS, ALEXANDER. *Introductory psychology: The principles* *of human adjustment.* New York: Rinehart, 1951. Pp. xv+461. $4.00.

Schneiders, Alexander. *Workbook in introductory psychology.* New York: Rinehart, 1951. Pp. 74. $1.25.

Segal, David. *State testing and evaluation programs.* Washington, D. C.: Federal Security Agency, Office of Education, 1951. (Circular No. 320.) Pp. 38.

Sibley, Elbrdige. *Support for independent scholarship and research.* New York: Social Science Research Council, 1951. Pp. xv+116. $1.25.

Smith, Glenn E. *Principles and practices of the guidance program: A basic text.* New York: Macmillan, 1951. Pp. xi+379. $3.25.

Stephens, J. M. *Educational psychology: The study of educational growth.* New York: Henry Holt, 1951. Pp. xxii+692. $4.25.

Stevens, S. S. (Ed.) *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. xi+1436. $15.00.

Strang, Ruth. *An introduction to child study.* (3rd ed.) New York: Macmillan, 1951. Pp. xi+705. $4.75.

Strecker, Edward A., Ebaugh, Franklin G., & Ewalt, Jack R. *Practical clinical psychiatry.* (7th ed.) Philadelphia: Blakiston, 1951. Pp. xii+506. $7.00.

Thurstone, L. L. *An experiment in the prediction of choice.* Chicago: Univ. of Chicago, Psychometric Laboratory, 1951. (No. 68.) Pp. 10.

Thurstone, L. L. *Experimental tests of temperament.* Chicago: Univ. of Chicago, Psychometric Laboratory, 1951. (No. 67.) Pp. 8.

Tilton, J. W. *An educational psychology of learning.* New York: Macmillan, 1951. Pp. vii+248. $3.50.

U. S. Children's Bureau. *Research relating to children: An inventory of studies in progress reported April 1950–July 1950.* Washington, D. C.: U. S. Children's Bureau, Federal Security Agency, 1951. (Suppl. No. 2.) Pp. 191.

Volkart, Edmund H. *Social behavior and personality: Contributions of W. I. Thomas to theory and social research.* New York: Social Science Research Council, 1951. Pp. ix+338. $3.00.

Welford, A. T. *Skill and age: An experimental approach.* New York: Oxford Univ. Press, 1951. Pp. ix+161. $1.75.

Zangwill, O. L. *An introduction to modern psychology.* New York: Philosophical Library, 1950. Pp. xi+227. $3.75.

Contributi del Laboratorio di Psicologia. *Pubbl. Univ. Cattolica Sacro Cuore,* 1950, Vol. 35. Pp. viii+386.

Permanent International Committee of Linguists. *Bibliographie linguistique de l'année 1948; et complement des années 1939–1947.* (Linguistic bibliography for the year 1948; and supplement for the years 1939–1947.) Utrecht-Brussels: Spectrum, 1951. Pp. xxiv+261. $4.50.

Symposia of the Society of Experimental Biology. (No. IV.) *Physiological mechanisms in animal behavior.* New York: Academic Press, 1951. $6.00.

INDEX OF SUBJECTS

Accident data, a chi-square test of proneness in, 496
Accident proneness, correlation vs curve fitting in research on, 413

Chi-square test for "proneness" in accident data, 496
Comments on Razran's discussion of stimulus generalization, 150
Comparison of a series of proportions: the problem and a new solution, 348
Computation of the level of significance in the F-test, 392
Concept formation, investigation of, 1
Concerning taste blindness to PTC, 419
Control group designs in human relations studies, 340
Correlation versus curve fitting in research on accident proneness: reply to Maritz, 413
Critical review of latent learning and related experiments, 97
Critical review of the validity and rationale of the forced-choice technique, 62

Daydreams, Freudian theory of (operational analysis), 369
Discussion group technique, history of leaderless, 383

Emotions conceptualized as intervening variables: with suggestions towards a theory of frustration, 465
Experimental evidence for the motivating function of reward, 130
Experimental naïveté and experiential naïveté, 327
Experiential naïveté, experimental naïveté and, 327

Forced-choice technique,
  critical review of validity and rationale of, 62
  note on Travers review, 435
  reply to Travers critical review of, 421
Freudian theory of daydreams; an operational analysis, 369
Frustration, suggestions towards a theory of, 465
F-test, computation of the level of significance in, 392

Genetics of schizophenia: reply to Pastore, 402

History of the leaderless discussion group technique, 383
Human relations studies, use of control group designs in, 340

"Individual Behavior," note on Louttit's review of, 79
Infrahuman mammals, sensory capacities of, 289
Intervening variables, emotions conceptualized as, 465
Investigation of concept formation, 1

Johnson's review of "Work and Effort," 75

Latent learning and related experiments, critical review of, 97
Latin squares, on use in psychology, 398
Leaderless discussion group technique, history of, 383
Letter to the Editor on Peters' reply to Lewis and Burke, 81
Level of significance in F-test, computation of, 392
Louttit's review of "Individual Behavior," a note on, 79

Mammals, sensory capacities of infrahuman, 289
Maritz, correlation vs curve-fitting in research on accident proneness: reply to, 413
Methodology in propaganda research, 32
Motivating function of reward, experimental evidence for, 130
Müller-Urban weights, psychophysical

method employing a modification of, 249

Naïveté, experimental and experiential, 327
Note on Louttit's review of "Individual Behavior," 79
Note on Travers critical review of forced-choice technique, 435

On the use of latin squares in psychology, 398
Operational analysis, Freudian theory of daydreams, 369

Pastore, reply to, on genetics of schizophrenia, 402
Peters' reply to Lewis and Burke, letter to the Editor on, 81
Phenylthiocarbamide and related compounds, taste blindness to, 71
pH, salivary, as a psychophysiological variable, 193
"Proneness" in accident data, a chi-square test of, 496
Propaganda research, methodology in, 32
Proportions, comparison of a series of, 348
Psychophysical method employing a modification of the Müller-Urban weights, 249
Psychophysiological variable, salivary pH as, 193
Psychological research, statistical consideration in, 156

Razran's discussion of stimulus generalization, comments on, 150
Reply, Ryan's "Work and Effort," 77
Reply to Travers' "Critical review of the validity and rationale of the forced-choice technique, 421
Research with the Wechsler-Bellevue Test: 1945–1950, 211
Reward, experimental evidence for the motivating function of, 130
Rorschach Test, some comments on the controlling of differences in responses on, 257
Ryan's "Work and Effort," a reply, 77

Salivary pH as a psychophysiological variable, 193
Schizophrenia, reply to Pastore on genetics of, 402
Sensory capacities of infrahuman mammals: vision, audition and gustation, 289
Skewness, three rational methods for reduction of, 505
Social psychology in World War II, studies in, 512
Some comments on the controlling of differences in responses on the Rorschach Test, 257
Statistical consideration in psychological research, 156
Stimulus generalization, comments on Razran's discussion of, 150
Studies in social psychology in World War II, 512

Taste blindness to phenylthiocarbamide and related compounds, 71
Taste blindness to PTC, 71
Three rational methods for the reduction of skewness, 505
Travers,
  note on (his) review of forced-choice technique, 435
  reply to (on forced-choice technique), 421

Use of extended control group designs in human relations studies, 340

Validity and rationale of the forced-choice technique, critical review of, 62

Wechsler-Bellevue test, research with (1945–1950), 211
"Work and Effort," Johnson's review of, 75

INDEX OF AUTHORS

GENERAL REVIEWS, SHORT ARTICLES, SPECIAL REVIEWS, NOTES

Ansbacher, H. L., 383
Ash, Philip, 289
Baier, D. E., 421
Bitterman, M. E., 75
Blum, M. L., 413
Boyd, W. C., 71
Brown, J. S., 465
Burke, C. J., 81, 392, 496
Burros, R. H., 505
Canter, R. R., Jr., 340
Christie, Richard, 327
Cohen, Jozef, 419
Creegan, R. F., 79
Davis, R. C., 348
Eichler, R. M., 257
Farber, I. E., 465
Furchtgott, Ernest, 193
Gilchrist, J. C., 193
Grice, G. R., 150
Guertin, W. H., 211
Harrison, Margaret J., 249
Harrison, Stephen, 249
Hurst, L. A., 402
Johnson, H. M., 77
Katz, Daniel, 512
McNemar, Quinn, 398
Mintz, Alexander, 413
Ogdon, D. P., 419
Rabin, A. I., 211
Razran, Gregory, 153
Richardson, M. W., 435
Seeman, William, 369
Seward, J. P., 130
Thistlethwaite, Donald, 97
Travers, R. M. W., 62
Vallance, T. R., 32
Vinacke, W. E., 1
Wilkinson, Bryan, 156

BOOKS REVIEWED

Abt, L. E., 539
Allport, G. W., 173
Anastasi, Anne, 265
Arnot, R., 183
Back, Kurt, 522
Bellak, Leopold, 539
Birch, H. G., 534
Bloom, Benjamin S., 533
Boring, E. G., 260
Bolles, M. Marjorie, 286
Brayfield, A. H., 538
Britt, Steuart Henderson, 448, 532
Broder, Lois J., 533
Brown, Warner, 366
Brožek, J., 530
Burgess, E. W., 85
Burrow, Trigant, 87
Cantril, Hadley, 447
Cavan, Ruth S., 85
Chapanis, A., 83
Chevigny, Hector, 182
Cochran, W. G., 272
Collins, Mary E., 523
Cottrell, L. S., Jr., 512
Cox, Gertrude M., 272
Deming, W. E., 453
Deutsch, Morton, 523
De Vinney, E. A., 512
Dorcus, R. M., 537
Edwards, A. L., 269
Engle, T. L., 263
Festinger, Leon, 522
Fleming, Charlotte M., 88
Fisher, V. E., 453
Foley, J. P., Jr., 265
Form, W. H., 527
Freeman, F. S., 279
Fromm-Reichmann, Frieda, 281
Fryer, D. H., 455
Garner, W. R., 83
Gibson, J. J., 360
Gilhousen, H. C., 366
Goldhamer, Herbert, 85
Gouldner, A. W., 445
Greenblatt, M., 183
Guilford, J. P., 93

Gulliksen, Harold, 458
Gurvitch, Georges, 163
Guthrie, E. R., 264
Guttman, L., 512
Hahn, M. E., 285
Harris, J. D., 273
Harris, R. E., 363
Harsh, C. M., 168
Hartley, E. L., 534
Hartley, Ruth E., 534
Havighurst, R. J., 85
Henry, E. R., 455
Henschel, A., 530
Hepner, H. W., 280
Homans, G. C., 449
Hoppock, Robert, 267
Horney, Karen, 542
Hurlock, Elizabeth B., 275
Hovland, C. I., 512
Janis, I. L., 512
Johns, R. L., 89
Jones, M. H., 537
Katz, David, 177
Kent, Grace H., 284
Keys, A., 530
Klineberg, Otto, 181
Kubie, L. S., 451
Landis, Carney, 286
Lefever, D. W., 283
Lewin, Kurt, 520
Lucas, Darrell Blaine, 532
Lumsdaine, A. A., 512
MacLean, M. S., 285
Maslow, A. H., 541
May, Rollo, 166
Mettler, F. A., 185
Mickelsen, O., 530
Miller, D. C., 527
Miller, J. G., 363
Mittelmann, Bela, 541
Morgan, C. T., 83
Mowrer, O. H., 438
Muench, G. A., 363
Munn, N. L., 175
Newcomb, T. M., 169
Parten, Mildred, 277
Pastore, Nicholas, 273
Plant, J. S., 268
Piaget, Jean, 535
Porteus, Stanley D., 529
Powers, F. F., 264
Prothro, E. T., 180
Rashevsky, Nicholas, 525
Redl, Fritz, 544
Roback, R. R., 164
Rosenblueth, Arturo, 444
Sargent, S. S., 161
Schacter, Stanley, 522
Scheinfeld, Amram, 456
Schrickel, H. G., 168
Sheffield, F. D., 512
Solomon, H. C., 183
Standen, Anthony, 90
Star, S., 512
Stone, C. P., 159
Stone, L. J., 363
Stouffer, S. A., 512
Suchman, E. A., 512
Taylor, H. L., 530
Teska, P. T., 180
Toman, Walter, 528
Tueber, H. L., 363
Turrell, Archie M., 283
Volgyesi, F. A., 442
von Bonin, Gerhardt, 276
von Frisch, Karl, 365
Watkins, J. G., 287
Wattenberg, W. W., 544
Wepman, J. M., 544
Wiener, Norbert, 367
Weitzel, H. I., 283
Williams, R. M., Jr., 512
Wortis, Joseph, 364
Zubin, J., 363
Zahl, P. A., 182

BOOK REVIEWERS

Anderson, J. E., 273
Beach, F. A., 365
Bellows, Carol S., 537
Bellows, R. M., 537
Berdie, R. F., 285
Black, Percy, 164
Boring, E. G., 177, 360
Burke, C. L., 525

Canter, R. R., Jr., 522
Carter, L. F., 163, 445
Cartwright, Dorwin, 181
Child, I. L., 269
Collier, R. M., 87
Courts, F. A., 263
Davis, F. B., 279
Davis, R. C., 366
Ellis, Albert, 542
Fischer, R. P., 286
Fitts, P. M., 83
Geldard, F. A., 273
Gilbert, W. M., 538
Graham, C. H., 260
Grant, D. A., 272
Gregory, W. S., 267
Guest, Lester P., 532
Hanna, J. V., 283
Harris, D. B., 265
Heathers, Glen, 169
Hemphill, J. K., 527
Herrington, L. P., 530
Hilgard, E. R., 438
Holtzman, W. H., 93, 535
Hunt, W. A., 363
Katz, Daniel, 512
Kendler, H. H., 159
Kuhlen, R. G., 85
Landis, Carney, 183
Lazar, Irving, 539
Littman, R. A., 268
London, I. D., 364
Louttit, C. M., 182
Marks, E. S., 277
McKinney, Fred, 544
McNemar, Quinn, 458
Miller, J. G., 451
Miller, Jessie L., 451
Morgan, C. T., 175
Mosteller, Frederick, 453
Mowrer, O. H., 281, 453
Muenzinger, K. F., 90, 528
Mussen, P. H., 541
Myklebust, H. R., 284
Nissen, H. W., 276
Norman, R. D., 287
O'Kelly, L. I., 544
Pepitone, Albert, 523
Price, F. B., 185
Rexroad, C. N., 180
Riggs, L. A., 444
Roback, A. A., 173
Roberts, S. O., 275
Rotter, J. B., 168
Ruja, Harry, 89
Scofield, C. F., 534
Shuttleworth, F. K., 264
Skinner, B. F., 367
Smith, M. B., 161, 449, 520
Sollenberger, R. T., 88
Stagner, Ross, 447
Strupp, H. H., 442
Terman, L. M., 456
Thorndike, R. L., 455
Vinacke, W. Edgar, 533
Walker, E. L., 438
Weschler, I. R., 280
Worcester, Dean A., 529
White, R. W., 166
Young, Kimball, 448

BUREAU OF EDNL. & PSYL. RESEARCH